



# When is selfish routing bad? The price of anarchy in light and heavy traffic

Riccardo Colini-Baldeschi, Roberto Cominetti, Panayotis Mertikopoulos,  
Marco Scarsini

## ► To cite this version:

Riccardo Colini-Baldeschi, Roberto Cominetti, Panayotis Mertikopoulos, Marco Scarsini. When is selfish routing bad? The price of anarchy in light and heavy traffic. *Operations Research*, 2020, 68 (2), pp.411-434. 10.1287/opre.2019.1894 . hal-01891558

**HAL Id: hal-01891558**

**<https://inria.hal.science/hal-01891558>**

Submitted on 9 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# WHEN IS SELFISH ROUTING BAD? THE PRICE OF ANARCHY IN LIGHT AND HEAVY TRAFFIC

RICCARDO COLINI-BALDESCHI\*, ROBERTO COMINETTI<sup>‡</sup>,  
PANAYOTIS MERTIKOPOULOS<sup>§</sup>, AND MARCO SCARSINI<sup>¶</sup>

**ABSTRACT.** This paper examines the behavior of the price of anarchy as a function of the traffic inflow in nonatomic congestion games with multiple origin-destination (O/D) pairs. Empirical studies in real-world networks show that the price of anarchy is close to 1 in both light and heavy traffic, thus raising the question: can these observations be justified theoretically? We first show that this is not always the case: the price of anarchy may remain a positive distance away from 1 for all values of the traffic inflow, even in simple three-link networks with a single O/D pair and smooth, convex costs. On the other hand, for a large class of cost functions (including all polynomials), the price of anarchy *does* converge to 1 in both heavy and light traffic, irrespective of the network topology and the number of O/D pairs in the network. We also examine the rate of convergence of the price of anarchy, and we show that it follows a power law whose degree can be computed explicitly when the network's cost functions are polynomials.

*“Traffic congestion is caused by vehicles, not by people in themselves.”*

— Jane Jacobs, *The Death and Life of Great American Cities*

## 1. INTRODUCTION

Almost every commuter in a major metropolitan area has experienced the frustration of being stuck in traffic. At best, this might mean being late for dinner; at worst, it means more accidents and altercations, not to mention the vastly increased damage to the environment caused by huge numbers of idling engines.

To name but an infamous example, China's G110 traffic jam in August 2010 brought to a standstill thousands of vehicles for 100 kilometers between Hebei and Inner Mongolia.

---

\* CORE DATA SCIENCE GROUP, FACEBOOK INC., 1 RATHBONE PLACE, LONDON, W1T 1FB, UK.

<sup>‡</sup> FACULTAD DE INGENIERÍA Y CIENCIAS, UNIVERSIDAD ADOLFO IBÁÑEZ, SANTIAGO, CHILE.

<sup>§</sup> UNIV. GRENoble ALPES, CNRS, INRIA, LIG, F-38000 GRENoble, FRANCE.

<sup>¶</sup> DIPARTIMENTO DI ECONOMIA E FINANZA, LUISS, VIALE ROMANIA 32, 00197 ROMA, ITALY.

*E-mail addresses:* [rickuz@fb.com](mailto:rickuz@fb.com), [roberto.cominetti@uai.cl](mailto:roberto.cominetti@uai.cl),  
[panayotis.mertikopoulos@imag.fr](mailto:panayotis.mertikopoulos@imag.fr), [marco.scarsini@luiss.it](mailto:marco.scarsini@luiss.it).

2010 *Mathematics Subject Classification.* Primary 91A13; secondary 91A43, 91A80.

*Key words and phrases.* Nonatomic congestion games; price of anarchy; light traffic; heavy traffic; regular variation.

We thank an anonymous reviewer of an earlier conference version of this paper for suggesting one of the examples with variable inflows in [Section 6](#).

This research benefited from the support of the FMJH Program PGM under grant HEAVY.NET and from the support of EDF, Thales, and Orange. R. Colini-Baldeschi and M. Scarsini are members of GNAMPA-INdAM. R. Cominetti and P. Mertikopoulos gratefully acknowledge the support and hospitality of LUISS during a visit in which this research was initiated. R. Cominetti's research is also supported by FONDECYT 1130564 and Núcleo Milenio ICM/FIC RC130003 “*Información y Coordinación en Redes*”. P. Mertikopoulos was partially supported by the ECOS/CONICYT Grant C15E03 and the Huawei HIRP Flagship project ULTRON. P. Mertikopoulos and M. Scarsini also gratefully acknowledge the support and hospitality of FONDECYT 1130564 and Núcleo Milenio “*Información y Coordinación en Redes*”.

The snarl-up lasted twelve days and resulted in drivers being unable to move for more than 1 kilometer per day, reportedly spending up to five days trapped in the jam. Not caused by weather or a natural disaster, this massive 10-day tie-up was instead laid at the feet of a bevy of trucks swarming on the shortest route to Beijing, thus clogging the highway to a halt (while ironically carrying supplies for construction work to ease congestion). This, therefore, raises the question: *how much better would things have been if all traffic had been routed by a social planner who could calculate (and enforce) the optimum traffic assignment?*

In game-theoretic terms, this question boils down to the inefficiency of Nash equilibria that are not Pareto optimal. The most widely used quantitative measure of this inefficiency is the so-called *price of anarchy* (PoA): introduced by Koutsoupias and Papadimitriou (1999) and so dubbed by Papadimitriou (2001), the price of anarchy is simply the ratio of the social cost of the least efficient Nash equilibrium divided by the minimum achievable social cost. By virtue of this straightforward definition, deriving worst-case bounds for the price of anarchy has given rise to a vigorous literature at the interface of operations research, economics and computer science, often leading to surprising and counter-intuitive results.

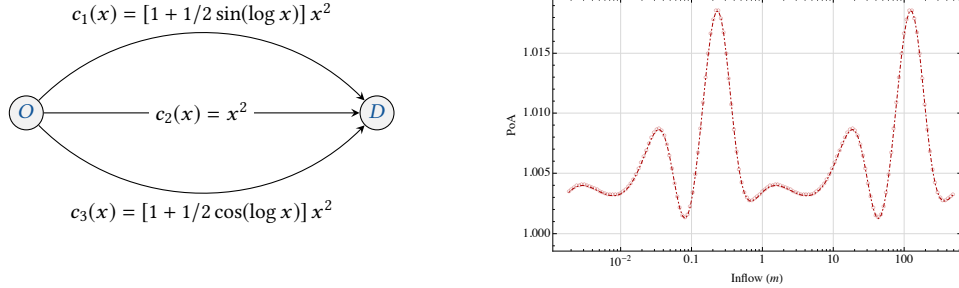
In the context of network congestion, Pigou (1920) was probably the first to note the inefficiency of selfish routing, and his elementary two-road example with a PoA of  $4/3$  is one of the two prototypical examples thereof. The other example is due to Braess (1968), and consists of a four-road network where the addition of a zero-cost segment makes things just as bad as in the Pigou case. These two examples were the starting point for Roughgarden and Tardos (2002) who showed that the price of anarchy in (nonatomic) routing games with affine costs may not exceed  $4/3$ . On the other hand, if the network's cost functions are polynomials of degree at most  $d$ , the price of anarchy may become as high as  $\Theta(d/\log d)$ , implying that selfish routing can become arbitrarily bad in networks with polynomial costs (Roughgarden, 2003).

By this token, and given the typically nonlinear relation between traffic loads and travel times, the intervention of a central planner seems necessary in order to regain some degree of efficiency. At the same time however, these worst-case instances are typically realized in networks with delicately tuned traffic loads and costs: if a network operates beyond this regime, it is not clear whether the price of anarchy remains high. In view of this, our aim in this paper is to examine the asymptotic behavior of the price of anarchy at both ends of the congestion spectrum: *light and heavy traffic*.

Using both analytical and numerical methods, a very recent study by O'Hare et al. (2016) suggests that the price of anarchy is usually close to 1 for both high and low traffic, and it fluctuates in the intermediate regime (typically exhibiting multiple local maxima). In a similar setting, Monnot et al. (2017) used a huge dataset on commuting students in Singapore to estimate the so-called "stress of catastrophe": this majorant of the ordinary price of anarchy was estimated to a value between 1.11 and 1.22, suggesting that the actual value of the price of anarchy in Singapore is lower (and definitely below the  $4/3$  worst-case bound of the Pigou/Braess examples).

All this leads to the following natural questions:

- a) Under what conditions does the price of anarchy converge to 1 in light/heavy traffic?
- b) Do these conditions depend on the network topology, its cost functions, or both?
- c) Can general results be obtained for networks with multiple origin-destination (O/D) pairs?



**Figure 1.** A network where selfish routing remains inefficient for both light and heavy traffic.

d) When these conditions are satisfied, how fast is this convergence?

**1.1. Our contributions.** Our first result is a cautionary tale: we show that the price of anarchy may oscillate between two bounds strictly greater than 1 for all values of the traffic inflow, even in simple parallel-link networks with a single O/D pair (cf. Fig. 1). The cost functions in our example are convex and differentiable, so neither convexity nor smoothness seem to play a major role in the efficiency of selfish routing. Moreover, our construction only involves a three-link network, so such phenomena may arise in any network containing such a three-link component.

Heuristically, the reason for this irregular – and, perhaps, counter-intuitive – behavior is that the growth rate of the network’s cost functions exhibits higher-order oscillations which persist at any scale, in both light and heavy traffic. To dispense with such pathologies, we focus on networks whose cost functions  $c_e(x)$  are asymptotically comparable to a *benchmark function*  $c(x)$  which is itself assumed to be *regularly varying* (cf. Definition 4.1). In so doing, we obtain a classification of the network’s edges, paths, and O/D pairs as *fast*, *slow* or *tight* relative to the chosen benchmark. Then, thanks to this classification, we obtain the following general result: *If the routing cost of the “most costly” O/D pair in the network behaves like the benchmark, the network’s price of anarchy converges to 1 in both light and heavy traffic.*

Polynomial cost functions satisfy all of the above requirements, leading to the comprehensive asymptotic principle:

*In networks with polynomial costs,  
the price of anarchy becomes 1 under both light and heavy traffic.*

In other words, a benevolent social planner with full control of traffic assignment would not do any better than selfish agents in conditions of high or low congestion. In particular, only if the traffic falls in an intermediate regime can there be a substantial gap between optimum and equilibrium states.

To assess how wide this intermediate regime might be in practice, we also examine the speed at which the price of anarchy converges to 1 as a function of the traffic inflow. Specializing to networks with polynomial costs, we find that in both regimes the convergence follows a power law with respect to the total traffic inflow, and we derive explicit sharp estimates for the corresponding rates.

**1.2. Related work.** Establishing worst-case bounds for the price of anarchy under different conditions has been a staple of the literature on congestion games ever since the

seminal result of Roughgarden and Tardos (2002). In words, this result states that in networks with affine costs the price of anarchy is no higher than  $4/3$ , independently of the network topology and/or the number of O/D pairs in the network. Furthermore, this bound is sharp in that, for every  $M > 0$ , there exists a network with traffic inflow  $M$  and affine costs such that the price of anarchy is *exactly* equal to  $4/3$ . Importantly, our analysis shows that the order of the quantifiers in the above statement *cannot be exchanged*: in any network with affine costs, the price of anarchy gets arbitrarily close to 1 if the traffic inflow is sufficiently large or small.

Worst-case bounds for the price of anarchy have been obtained for larger classes of cost functions. For polynomial costs with degree at most  $d$ , Roughgarden (2003) showed that the worst possible instance grows as  $\Theta(d/\log d)$  while Dumrauf and Gairing (2006) provided sharper bounds for monomials of maximum degree  $d$  and minimum degree  $q$ . Extending the above results, Roughgarden and Tardos (2004) provided a unifying result for costs that are differentiable with  $xc(x)$  convex, while Correa et al. (2004, 2008) considered less regular classes of cost functions. Correa et al. (2007) also studied the price of anarchy when the goal is to minimize the maximum – rather than the average – latency in the network. For a survey, the reader is referred to Roughgarden (2007).

In a more practical setting, Youn et al. (2008) studied the difference between optimal and actual system performance in real transportation networks, focusing in particular on Boston’s road network. They observed that the price of anarchy depends crucially on the total traffic inflow: it starts at 1, it then grows with some oscillations, and ultimately returns to 1 as the flow increases. González Vayá et al. (2015) studied optimal scheduling for the electricity demand of a fleet of plug-in electric vehicles: without using the term, they showed that the price of anarchy goes to 1 as the number of vehicles grows. Cole and Tao (2016) showed that in large Walrasian auctions and in large Fisher markets the price of anarchy goes to one as the market size increases. Finally, Feldman et al. (2016) took a different asymptotic approach and considered atomic games where the number of players grows to infinity. Applying the notion of  $(\lambda, \mu)$ -smoothness to the resulting sequence of atomic games, they showed that the price of anarchy converges to the corresponding nonatomic limit.

From an analytic standpoint, the closest antecedent to our paper is the recent work of Colini-Baldeschi et al. (2016) who studied the heavy traffic limit of the price of anarchy in parallel networks with a single O/D pair. Their analysis identified that regular variation plays an important role in this setting; however, it offered no insights into non-parallel networks with multiple O/D pairs or the light traffic regime. Our paper provides an in-depth answer to these questions: we show that (a) regular variation yields asymptotic efficiency under both light and heavy traffic conditions; (b) the topology of the network doesn’t matter; and (c) the existence of several O/D pairs doesn’t matter as long as they admit a common benchmark (which is always the case if the network’s cost functions are polynomial).

Building on a previous unpublished version of the present paper, Wu et al. (2017) introduced a class of congestion games, called *scalable*, whose price of anarchy converges to 1 as the total demand diverges. They also computed the rate of convergence of the price of anarchy for the special case of BPR cost functions of the same degree. Stidham (2014) also studied the behavior of the price of anarchy for queueing networks in heavy traffic under various assumptions on the structure of the network and on the stochastic properties of the queues. His results can be used for the analysis of network routing models with capacity constraints.

Our work should also be compared to that of Monnot et al. (2017) who performed an empirical study of the price of anarchy based on data from thousands of commuting students in Singapore. Focusing on the network’s “stress of catastrophe” (an empirical majorant of the network’s price of anarchy), they showed that routing choices are near-optimal and the incurred price of anarchy is much lower than what traditional worst-case bounds suggest. Interestingly, the study of Monnot et al. (2017) also suggests that the Singapore road network is often lightly congested: as such, their results can be seen as a practical validation of the light traffic results presented here (and, conversely, our results provide a theoretical justification for their empirical observations).

**1.3. Outline of the paper.** The paper is organized as follows. In Section 2, we introduce the basic model and concepts that will be used in the rest of the paper. Section 3 provides two motivating examples for the analysis to follow. In Section 4, we treat networks with a single O/D pair, whereas Section 5 examines networks with multiple O/D pairs. The more complicated case of variable relative inflows is treated in Section 6. Finally, in Section 7, we study the rate of convergence of the price of anarchy in light and heavy traffic. To streamline our presentation, the proofs of our main results have been relegated to a series of appendices at the end of the paper.

## 2. MODEL AND PRELIMINARIES

**2.1. Network model.** Following Beckmann et al. (1956) and Roughgarden and Tardos (2002), the basic component of our model will be a directed multi-graph  $\mathcal{G} \equiv \mathcal{G}(\mathcal{V}, \mathcal{E})$  with vertex set  $\mathcal{V}$  and edge set  $\mathcal{E}$  (both finite). We further assume that there is a finite set of *origin-destination* (O/D) pairs indexed by  $i \in \mathcal{I}$ , each with an individual *traffic demand*  $m^i \geq 0$  that is to be routed from the pair’s *origin node*  $O^i \in \mathcal{V}$  to its *destination*  $D^i \in \mathcal{V}$ . To route this traffic, the  $i$ -th O/D pair employs a set  $\mathcal{P}^i$  of *paths* joining  $O^i$  to  $D^i$ , with each path  $p \in \mathcal{P}^i$  comprising a sequence of edges that meet head-to-tail in the usual way; specifically, we do not assume that  $\mathcal{P}^i$  is necessarily the set of *all* paths joining  $O^i$  to  $D^i$ , but only some subset thereof.<sup>1</sup> For bookkeeping reasons, we will also make the following standing assumptions throughout our paper:

- a) The *total inflow rate*  $M = \sum_{i \in \mathcal{I}} m^i$  is positive (so there is a nonzero amount of traffic in the network).
- b) The path sets  $\mathcal{P}^i$  are disjoint (which in particular holds trivially if all pairs  $(O^i, D^i)$  are distinct).

Now, writing  $\mathcal{P} \equiv \bigcup_{i \in \mathcal{I}} \mathcal{P}^i$  for the union of all such paths, the set of feasible *routing flows*  $f = (f_p)_{p \in \mathcal{P}}$  in the network is defined as

$$\mathcal{F} = \{f \in \mathbb{R}_+^{\mathcal{P}} : \sum_{p \in \mathcal{P}^i} f_p = m^i \text{ for all } i \in \mathcal{I}\}. \quad (2.1)$$

In turn, a routing flow  $f \in \mathcal{F}$  induces a *load* on each edge  $e \in \mathcal{E}$  as

$$x_e = \sum_{p \ni e} f_p, \quad (2.2)$$

and we write  $x = (x_e)_{e \in \mathcal{E}}$  for the corresponding *load profile* on the network. Given all this, the delay (or latency) experienced by an infinitesimal traffic element traversing edge  $e$  is determined by a nondecreasing continuous *cost function*  $c_e : [0, \infty) \rightarrow [0, \infty)$ : more precisely, if  $x = (x_e)_{e \in \mathcal{E}}$  is the load profile induced by a feasible routing flow  $f = (f_p)_{p \in \mathcal{P}}$ ,

<sup>1</sup>This distinction is particularly relevant for packet-switched networks (such as the Internet) where only paths with a low hop count are typically employed.

the incurred delay on edge  $e \in \mathcal{E}$  is  $c_e(x_e)$ . Hence, with a slight abuse of notation, the associated cost of path  $p \in \mathcal{P}$  will be given by

$$c_p(f) = \sum_{e \in p} c_e(x_e). \quad (2.3)$$

Putting together all of the above, the tuple  $\Gamma = (\mathcal{G}, \mathcal{I}, \{m^i\}_{i \in \mathcal{I}}, \{\mathcal{P}^i\}_{i \in \mathcal{I}}, \{c_e\}_{e \in \mathcal{E}})$  will be referred to as a (*nonatomic*) *routing game*. When we want to explicitly keep track of the total inflow rate  $M = \sum_i m^i$ , we write  $\Gamma_M$  instead of  $\Gamma$ ; also, when there is a single O/D pair, we will drop all reference to  $i$  and  $\mathcal{I}$  altogether.

**2.2. Equilibrium, optimality, and the price of anarchy.** In routing games, the notion of Nash equilibrium is captured by Wardrop's first principle: *at equilibrium, the delays along utilized paths are equal and no higher than those that would be experienced by an infinitesimal traffic element going through an unused route* (Wardrop, 1952).

Formally, a routing flow  $f^*$  is said to be a *Wardrop equilibrium* (WE) of  $\Gamma$  if, for all  $i \in \mathcal{I}$ , we have

$$c_p(f^*) \leq c_{p'}(f^*) \quad \text{for all } p, p' \in \mathcal{P}^i \text{ such that } f_p^* > 0. \quad (2.4)$$

From the work of Beckmann et al. (1956), it is known that Wardrop equilibria coincide with the solutions of the (convex) minimization problem

$$\begin{aligned} & \text{minimize} && \sum_{e \in \mathcal{E}} C_e(x_e), \\ & \text{subject to} && x_e = \sum_{p \ni e} f_p, \quad f \in \mathcal{F}, \end{aligned} \quad (\text{WE})$$

where  $C_e(x_e) = \int_0^{x_e} c_e(w) dw$  denotes the primitive of  $c_e$ . Analogously, *socially optimum* (SO) flows are defined as solutions of the total cost minimization problem

$$\begin{aligned} & \text{minimize} && \sum_{p \in \mathcal{P}} f_p c_p(f_p), \\ & \text{subject to} && f \in \mathcal{F}. \end{aligned} \quad (\text{SO})$$

By a simple rearrangement of terms, the objective function of (SO) can be rewritten as  $L(x) = \sum_{e \in \mathcal{E}} x_e c_e(x_e)$ , so the value of the above problem can be expressed equivalently (and somewhat more concisely) as

$$\text{Opt}(\Gamma) = \min_{x \in \mathcal{X}} L(x), \quad (2.5)$$

where  $\mathcal{X} = \{x \in \mathbb{R}_+^{\mathcal{E}} : x_e = \sum_{p \ni e} f_p, f \in \mathcal{F}\}$  denotes the set of all load profiles of the form (2.2). Thus, to quantify the gap between solutions to (WE) and (SO), let

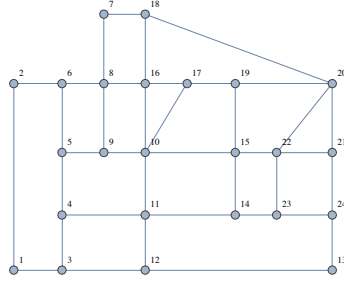
$$\text{Eq}(\Gamma) = L(x^*) \quad (2.6)$$

where  $x^*$  is the load profile induced by a Wardrop equilibrium  $f^*$  of  $\Gamma$  (by a standard result of Beckmann et al. (1956), all such flows incur the same total cost). The game's *price of anarchy* (PoA) is then defined as

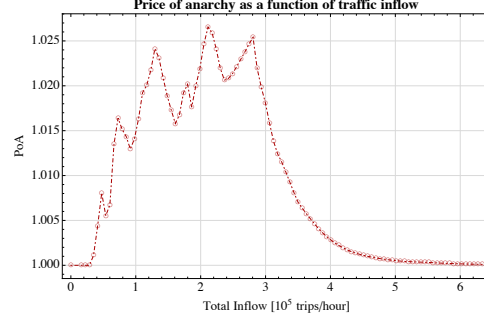
$$\text{PoA}(\Gamma) = \frac{\text{Eq}(\Gamma)}{\text{Opt}(\Gamma)}. \quad (2.7)$$

For this ratio to be well-defined, we must have  $\text{Opt}(\Gamma) > 0$ ; otherwise, if this is not the case, we will vacuously set  $\text{PoA}(\Gamma) = 1$ . To avoid such technicalities, we will tacitly assume that  $\text{Opt}(\Gamma) > 0$  throughout.





(a) The Sioux Falls road network.



(b) Anarchy and efficiency in Sioux Falls.

**Figure 2.** The price of anarchy in the Sioux Falls metropolitan area as a function of the traffic inflow.

Of course,  $\text{PoA}(\Gamma) \geq 1$  with equality if and only if Wardrop equilibria are also socially efficient. Our main objective in what follows will be to study the asymptotic behavior of this ratio when  $M \rightarrow 0$  or  $M \rightarrow \infty$ .

### 3. FIRST RESULTS

**3.1. Sioux Falls: a representative case study.** To motivate our analysis, we begin by examining the behavior of the price of anarchy in the road network of Sioux Falls, a standard case study in the transportation literature. For concreteness, the network's (two-way) arterial roads are shown in Fig. 2(a) and their delay functions are taken to be of the BPR (Bureau of Public Roads) type

$$c_e(x) = a_e + b_e x^{d_e} \quad (3.1)$$

with coefficients  $a_e, b_e$  and degrees  $d_e$  (typically  $d_e = 4$ ) taken from the standard reference work of LeBlanc et al. (1975, Table 1). To analyze the network's price of anarchy as a function of the total traffic inflow, we considered all 528 O/D pairs with nonzero inflow, and for each O/D pair  $i \in \mathcal{I}$ , we restricted  $\mathcal{P}_i$  to contain only the five shortest paths in terms of free-flow travel time (i.e., the time taken to traverse a path when empty). We then scaled up or down these inflows preserving the ratios between different O/D pairs and we plotted the network's price of anarchy for various values of the total inflow  $M$ .

As can be seen from Fig. 2(b), the network's price of anarchy is identically equal to 1 when the total inflow is small enough (approximately up to  $3.7 \times 10^4$  trips per hour); for intermediate values of  $M$ , the price of anarchy becomes strictly greater than 1, and, ultimately, it decreases monotonically to 1 in the heavy traffic limit. Interestingly, LeBlanc et al. (1975) report a value of  $M_{\text{avg}} \approx 3.6 \times 10^5$  trips/hour for the network's median traffic inflow; this value is well within the range where the price of anarchy decreases monotonically to 1 and, indeed, the observed value is approximately equal to 1.005, indicating a 0.5% difference between socially optimum and equilibrium flows under median traffic conditions.

Similar conclusions have been drawn in the literature from empirical studies in London, New York and Boston (Youn et al., 2008), as well as Sioux Falls with different subsets of O/D pairs and connecting paths per pair (O'Hare et al., 2016). In particular, in all cases, it was observed that there is an initial interval of values of  $M$  for which the price of anarchy is identically equal to 1; our first result shows that this is not mere happenstance:



**Proposition 3.1.** *If the network's cost functions are of the form (3.1) with  $a_e, b_e > 0$  and  $d_e \equiv d$  for all  $e \in \mathcal{E}$ , we have  $\text{PoA}(\Gamma_M) = 1$  for all sufficiently small  $M$ .*

In fact, as the following result shows, this behavior arises whenever each O/D pair admits a single “best” path under zero inflow:

**Proposition 3.2.** *Let  $\mathcal{P}_{\min}^i = \arg \min_{p \in \mathcal{P}^i} c_p(0)$  denote the set of minimal cost paths of the  $i$ -th O/D pair under zero inflow. If  $\mathcal{P}_{\min}^i$  is a singleton for all  $i \in \mathcal{I}$ , we have  $\text{PoA}(\Gamma_M) = 1$  for all sufficiently small  $M$ .*

The above results (both proven in [Appendix A](#)) provide a reasonable theoretical justification for the light traffic behavior of the price of anarchy that is observed in [Fig. 2](#) (the heavy traffic limit is discussed in detail in the next sections). At the same time however, the BPR and “unique best path” assumptions in [Propositions 3.1](#) and [3.2](#) respectively suggest that there is a finer mechanism at play which becomes apparent when the total cost at low traffic depends more delicately on the distribution of traffic in the network. We make this precise in the following section where we provide an example of a three-link network where the price of anarchy oscillates between two values strictly greater than 1, for all values of the traffic inflow.

**3.2. A network where selfish routing is always inefficient.** To construct an example of an “always inefficient” network, our approach will be to take a network with a certain degree of periodicity, obtain an explicit handle for its price of anarchy over a compact interval, and then tessellate this behavior over the entire traffic spectrum  $(0, \infty)$ . To carry this out, let  $\Gamma_M$  be a nonatomic routing game consisting of a single O/D pair with traffic inflow  $M$ . This traffic is to be routed over the three-link parallel graph of [Fig. 1](#) with cost functions

$$c_1(x_1) = x_1^d \left[ 1 + \frac{1}{2} \sin(\log x_1) \right], \quad (3.2a)$$

$$c_2(x_2) = x_2^d, \quad (3.2b)$$

$$c_3(x_3) = x_3^d \left[ 1 + \frac{1}{2} \cos(\log x_3) \right], \quad (3.2c)$$

where  $d$  is a positive integer. It is easy to see that these cost functions are increasing, strictly convex and smooth on  $[0, \infty)$  for all  $d \geq 2$ , and they all grow as  $\Theta(x^d)$  at both traffic limits ( $x \rightarrow 0$  and  $x \rightarrow \infty$ ). Furthermore, the functions  $x_e c_e(x_e)$  are *strictly* convex, so the optimum traffic allocation problem (SO) admits a unique solution. Hence, the only way for the game's price of anarchy to be equal to 1 is when the game's (also unique) Wardrop equilibrium coincides with the network's socially optimum flow.

As we show in [Appendix A](#), the equations determining the network's equilibrium and optimum flows never admit a common solution, so the price of anarchy is strictly greater than 1 over any compact interval (showing in this way that the conclusion of [Propositions 3.1](#) and [3.2](#) already fails for this example). Moreover, the trigonometric terms in (3.2) imply that these equations are periodic in a logarithmic scale (i.e., in  $\log M$ ). Hence, combining these two properties, we obtain:

**Proposition 3.3.** *In the three-link parallel network defined above,  $\text{PoA}(\Gamma_M)$  is periodic in  $\log M$  and oscillates between two values strictly greater than 1.*

[Colini-Baldeschi et al. \(2016\)](#) already provided examples of networks where  $\limsup_{M \rightarrow \infty} \text{PoA}(\Gamma_M) > 1$  but the cost functions involved were fairly irregular and the  $\liminf$  of the price of anarchy was still 1 (i.e., selfish routing was still efficient infinitely often). By contrast, in the above example, the price of anarchy is bounded away from 1 for *all* possible demands,

and this despite the fact that the network's cost functions are smooth, strictly convex and grow as  $\Theta(x^d)$  at both ends of the congestion spectrum. This is a considerable sharpening of the example of Colini-Baldeschi et al. (2016) as it shows that there are cases where efficiency is *never* achieved at equilibrium – not even asymptotically.

#### 4. NETWORKS WITH A SINGLE O/D PAIR

Despite the highly smooth and convex structure of the example network of Proposition 3.3, closer inspection reveals that the growth rate of its cost functions exhibits persistent oscillations at both 0 and  $\infty$ . This naturally leads to the following question: *Does selfish routing remain bad for “reasonable” cost functions that do not behave irregularly in the limit?*

To quantify – and discard – such irregularities, we will employ the seminal notion of *regular variation* (recalled below). For clarity and concision, we will focus for now on networks with a single O/D pair; the case of multiple O/D pairs will be discussed in detail later, in Sections 5 and 6.

**4.1. Regular variation and edge classification.** To present a unified perspective, we will tackle both ends of the congestion spectrum simultaneously by introducing the *traffic limit indicator*  $\omega \in \{0, \infty\}$ : letting  $M \rightarrow \omega$  gives the *light traffic limit* for  $\omega = 0$  and the *heavy traffic limit* for  $\omega = \infty$ . Regular variation at either limit is then defined as follows:

**Definition 4.1.** A function  $g: (0, \infty) \rightarrow (0, \infty)$  is said to be *regularly varying at  $\omega$*  if

$$\lim_{t \rightarrow \omega} \frac{g(tx)}{g(t)} \text{ is finite and nonzero for all } x > 0. \quad (4.1)$$

In words, regular variation means that  $g(t)$  grows at the same rate when viewed at different scales (determined here by  $x$ ). The concept itself dates back to Karamata (1930, 1933) and has been used extensively in functional analysis, probability, and large deviations theory (see e.g., de Haan and Ferreira, 2006; Jessen and Mikosch, 2006; Resnick, 2007); for a comprehensive survey we refer the reader to Bingham et al. (1989).

Standard examples of regularly varying functions include all affine functions, polynomials, logarithms, and, more generally, all analytic functions (barring those with an essential singularity at  $\omega = \infty$ ).<sup>2</sup> On the other hand, despite being bounded from above and below as  $\Theta(x^d)$ , the oscillatory cost functions (3.2) used in the counterexample of Section 3 are *not* regularly varying. Indeed, at either  $\omega = 0$  or  $\omega = \infty$ , the limit

$$\lim_{t \rightarrow \omega} \frac{c_1(tx)}{c_1(t)} = \lim_{t \rightarrow \omega} \frac{1 + \frac{1}{2} \sin(\log t + \log x)}{1 + \frac{1}{2} \sin(\log t)} x^d \quad (4.2)$$

does not exist in  $(0, \infty)$  unless  $\log x = k\pi$  for some  $k \in \mathbb{Z}$  (and likewise for  $c_3$ ). In this way, regular variation provides a much finer view than polynomial growth.

With all this at hand, we will dispose of growth irregularities like the above by positing that the network's cost functions can be compared asymptotically to some regularly varying function  $c(x)$ . Specifically, given an ensemble of cost functions  $\mathcal{C} = \{c_e\}_{e \in \mathcal{E}}$ , we will say that a regularly varying function  $c: (0, \infty) \rightarrow (0, \infty)$  is a *benchmark for  $\mathcal{C}$  at  $\omega$*  if the following (possibly infinite) limit exists for all  $e \in \mathcal{E}$

$$\alpha_e = \lim_{x \rightarrow \omega} \frac{c_e(x)}{c(x)}. \quad (4.3)$$

<sup>2</sup>Recall here that a function  $g(x)$  is analytic on a domain  $U$  if it is equal to its Taylor series on  $U$ .

This limit will be called the *index of edge  $e$  at  $\omega$* , and  $e$  will be called *fast*, *slow*, or *tight* (relative to  $c$  at  $\omega$ ) if  $\alpha_e$  is respectively 0,  $\infty$ , or in-between. In particular, when  $e$  is tight,  $c_e(x)$  is also regularly varying and exhibits the same asymptotic behavior as the benchmark function  $c(x)$  at  $\omega$ ; if  $e$  is fast, then  $c_e(x) = o(c(x))$ ; and, finally, if  $e$  is slow, then  $c(x) = o(c_e(x))$ . As such, a benchmark function groups the network's edges into three equivalence classes that exhibit the same qualitative behavior with respect to  $c(x)$ .

Of course, this partition depends on the chosen benchmark and the traffic limit (light or heavy): for instance,  $x^2$  is fast with respect to  $x$  at 0, but it is slow at  $\infty$ . For concision, we will not keep track of this dependence explicitly and instead rely on the context to resolve any ambiguities. However, it will be important to keep in mind that the classification of fast and slow edges could be flipped when transitioning from heavy to light traffic and vice versa.

Now, since bottlenecks along a path are caused by its slowest edges, we also define the *index of a path  $p \in \mathcal{P}$*  as

$$\alpha_p = \max_{e \in p} \alpha_e, \quad (4.4)$$

and we say that  $p$  is *fast*, *slow*, or *tight* based on whether  $\alpha_p$  is 0,  $\infty$ , or in-between. Finally, given that traffic will tend to be routed along the fastest paths in the network, we define the *index of the network* as

$$\alpha = \min_{p \in \mathcal{P}} \alpha_p, \quad (4.5)$$

and we say that the network is itself *tight* if  $0 < \alpha < \infty$ . In words, a path is fast (resp. tight/slow) if its slowest edge is fast (resp. tight/slow), and a network is tight if its fastest path is tight.

Defined this way, tightness guarantees that the network admits a path whose cost behaves asymptotically as a (positive) multiple of the benchmark function  $c(x)$ . The importance of this requirement is again illustrated by the cost model (3.2) of the previous section: if we only assumed that the network admits a path whose cost behaves as  $\Theta(c(x))$ , then we would not be able to rule out the pathological oscillations of the example in Section 3.

**4.2. The light traffic limit.** Thanks to the above legwork, we are in a position to state our main result for lightly congested networks with a single O/D pair:

**Theorem 4.2.** *Let  $\Gamma_M$  be a nonatomic routing game with a single O/D pair. If the network is tight under light traffic ( $\omega = 0$ ), then*

$$\lim_{M \rightarrow 0} \text{PoA}(\Gamma_M) = 1. \quad (4.6)$$

In words, Theorem 4.2 simply states that if the cost of the network's fastest path is regularly varying at 0, selfish routing becomes efficient in light traffic. To streamline our presentation, Theorem 4.2 is proved in Appendix B as a special case of a much more general statement. Here, we focus on some immediate corollaries thereof:

**Corollary 4.3.** *Suppose that, for every edge  $e \in \mathcal{E}$ , the limit  $\lim_{x \rightarrow 0} c_e(x)/x^{q_e}$  is finite and nonzero for some  $q_e \geq 0$ . Then,  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow 0$ .*

*Proof.* Referring to  $q_e$  as the *order* of  $e$ , define the order of a path  $p \in \mathcal{P}$  as  $q_p = \min_{e \in p} q_e$  and that of the network as  $q = \max_{p \in \mathcal{P}} q_p$ . Clearly,  $\lim_{x \rightarrow 0} c_e(x)/x^q = 0$  if and only if  $q_e > q$ ;  $\lim_{x \rightarrow 0} c_e(x)/x^q = \infty$  if and only if  $q_e < q$ ; and  $\lim_{x \rightarrow 0} c_e(x)/x^q \in (0, \infty)$  if and only if  $q_e = q$ . This shows that the network is tight with respect to  $c(x) = x^q$  at 0, so Theorem 4.2 applies.  $\square$

**Corollary 4.4.** *In a single O/D-network with analytic costs we have  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow 0$ .*

*Proof.* If  $c_e(x) = \sum_{k=0}^{\infty} c_{k,e} x^k$  for small enough  $x$ , we have  $\lim_{x \rightarrow \infty} c_e(x)/x^{q_e} \in (0, \infty)$  for  $q_e = \min\{k \in \mathbb{N} : c_{k,e} \neq 0\}$ . Our claim then follows from [Corollary 4.3](#).  $\square$

**Corollary 4.5.** *In a single O/D-network with polynomial costs we have  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow 0$ .*

Of the above results, [Corollaries 4.4](#) and [4.5](#) are of special practical interest because most latency models that have been proposed in the literature are polynomial or analytic at 0. In urban networks, the golden standard is the Bureau of Public Roads (BPR) quartic model  $c_e(x) = a_e + b_e x^4$ , while basically all of the established queueing models used in the theory of packet-switched networks ( $M/M/1$ ,  $M/G/k$ ,  $M/M/c$ , etc.) are analytic at 0 ([Bertsekas and Gallager, 1992](#)).

Despite appearances, the very wide applicability of [Theorem 4.2](#) and its corollaries is fairly surprising. Indeed, at first sight, one would expect that when  $M \rightarrow 0$ , traffic is so light that it doesn't really matter how it is routed. This is indeed the case if, for instance, all paths in the network exhibit different positive costs when  $M = 0$  (cf. [Proposition 3.2](#)). However, if the cost of an empty path is zero, this is no longer the case: the optimum and equilibrium traffic assignments could be fairly different (even when the network is lightly congested), so there is no a priori reason for the price of anarchy to converge to 1 as  $M \rightarrow 0$  (the example of Section 3 clearly illustrates this phenomenon). [Theorem 4.2](#) shows that all that is needed for this to occur is for the network's cost functions to be faithfully represented by a common benchmark function: when this condition is met, optimum and equilibrium costs no longer fluctuate but, instead, they converge to the same value.

**4.3. The heavy traffic limit.** Our main result for highly congested networks with a single O/D pair is as follows:

**Theorem 4.6.** *Let  $\Gamma_M$  be a nonatomic routing game with a single O/D pair. If the network is tight under heavy traffic ( $\omega = \infty$ ), then*

$$\lim_{M \rightarrow \infty} \text{PoA}(\Gamma_M) = 1. \quad (4.7)$$

In words, [Theorem 4.6](#) simply states that if the cost of the network's fastest path is regularly varying at  $\infty$ , selfish routing becomes efficient in heavy traffic. To compare and contrast the light and heavy traffic regimes, we relegate the proof of [Theorem 4.6](#) to [Appendix B](#) and only focus here on some immediate corollaries thereof:

**Corollary 4.7.** *Suppose there exists a path  $p \in \mathcal{P}$  with bounded costs, that is,  $\lim_{x \rightarrow \infty} c_e(x) < \infty$  for all  $e \in p$ . Then,  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \infty$ .*

*Proof.* Taking  $c(x) = 1$ , we get  $\alpha_e = \lim_{x \rightarrow \infty} c_e(x) \in (0, \infty]$  for all  $e \in \mathcal{E}$ . By assumption, there exists a path such that  $0 < \alpha_p < \infty$ , so [Theorem 4.6](#) applies.  $\square$

**Corollary 4.8.** *Suppose that the limit  $\lim_{x \rightarrow \infty} c_e(x)/x^{q_e}$  is finite and nonzero for some  $q_e \geq 0$  and all  $e \in \mathcal{E}$ . Then,  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \infty$ .*

*Proof.* Shadowing the proof of [Corollary 4.3](#), let  $q_p = \max_{e \in p} q_e$  and  $q = \min_{p \in \mathcal{P}} q_p$  (but note the reversal of the max and min operators). Clearly,  $\lim_{x \rightarrow \infty} c_e(x)/x^q = 0$  if and only if  $q_e < q$ ;  $\lim_{x \rightarrow \infty} c_e(x)/x^q = \infty$  if and only if  $q_e > q$ ; finally,  $\lim_{x \rightarrow \infty} c_e(x)/x^q \in (0, \infty)$  if and only if  $q_e = q$ . This shows that the network is tight with respect to  $c(x) = x^q$  at  $\infty$ , so [Theorem 4.6](#) applies.  $\square$

**Corollary 4.9.** *In a single O/D-network with polynomial costs we have  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \infty$ .*

In a certain, precise sense, [Theorems 4.2](#) and [4.6](#) show that the high and low congestion regimes can be seen as different sides of the same coin. By excluding pathological oscillations at either end of the congestion spectrum, regular variation ensures asymptotic regularity and guarantees that selfish routing becomes efficient in the limit: specifically, tightness at 0 guarantees efficiency in light traffic while tightness at  $\infty$  guarantees efficiency in heavy traffic. By this token, taking [Corollaries 4.5](#) and [4.9](#) in tandem implies that selfish routing becomes efficient under both light and heavy traffic in networks with polynomial costs and a single O/D pair.

That being said, there are still important, quantitative differences between the light and heavy traffic limits. For instance, even though [Corollaries 4.8](#) and [4.9](#) are direct analogues of their light traffic counterparts, the conclusion of [Corollary 4.7](#) is false in light traffic (the three-link network of [Section 3](#) serves again as a counterexample). In fact, even in the case of polynomial costs ([Corollary 4.5](#) vs. [Corollary 4.9](#)), there is an important reversal of roles that takes place between fast and slow edges. Specifically, edges that are fast in light traffic typically become slow under heavy traffic and vice versa (importantly however, tight edges are not re-classified under this regime change). Nevertheless, despite this reversal, the price of anarchy still goes to 1 in both cases.

## 5. NETWORKS WITH MULTIPLE O/D PAIRS

We now extend our analysis to networks with multiple O/D pairs. In this case, if the inflow rate of the  $i$ -th O/D pair is  $m^i$ , the total traffic inflow in the network is given by

$$M = \sum_{i \in \mathcal{I}} m^i, \quad (5.1)$$

and we write

$$\lambda^i = \frac{m^i}{M} \quad (5.2)$$

for the *relative inflow rate* of the  $i$ -th O/D pair – i.e., the fraction of the total traffic generated by the pair in question. In the rest of this section, we will assume that the relative inflow of every O/D pair  $i \in \mathcal{I}$  is a fixed positive constant that does not depend on  $M$ ; the case of variable inflow rates will be discussed in detail in [Section 6](#).

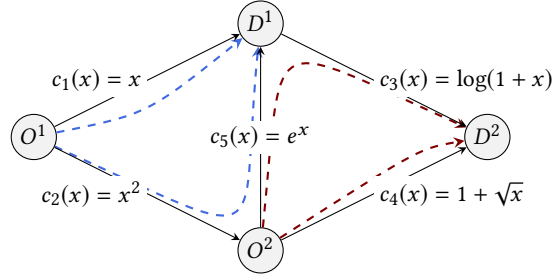
The key difference with the single-pair setting is that routing costs for different O/D pairs may exhibit completely different asymptotic behaviors in the limit. As a result, in the presence of multiple O/D pairs, the definition of the network’s index (and the related notion of tightness) must be re-examined. To do so, given that the traffic generated by an O/D pair will tend to be routed along the pair’s fastest path, we first define the *index of an O/D pair*  $i \in \mathcal{I}$  as

$$\alpha^i = \min_{p \in \mathcal{P}^i} \alpha_p. \quad (5.3)$$

Just like edges and paths, this index can be used to classify O/D pairs as *fast*, *slow* or *tight* depending on whether  $\alpha^i$  is respectively 0,  $\infty$ , or in-between. The *index of the network* is then defined as

$$\alpha = \max_{i \in \mathcal{I}} \alpha^i, \quad (5.4)$$

and we say that the network is *tight* if  $0 < \alpha < \infty$ . Heuristically, this definition simply captures the fact that the leading contribution to congestion is due to the “costliest” O/D



**Figure 3.** A Wheatstone network with two O/D pairs (cf. Example 5.1 below). In heavy traffic, the network is tight relative to the benchmark function  $c(x) = x$ ; in light traffic, the network is tight relative to the benchmark  $c(x) = 1$ .

pairs in the network; obviously, if there is but a single O/D pair, this last point is moot and (5.4) reduces to (4.5).

*Example 5.1.* To illustrate the above concepts, consider a Wheatstone network with two O/D pairs and cost functions as in Fig. 3. Focusing first on the heavy traffic limit, the benchmark  $c(x) = x$  would classify edge 1 as tight, edges 2 and 5 as slow, and edges 3 and 4 as fast. Accordingly, the first O/D pair would be classified as tight while the second O/D pair would be classified as fast; since no pair is slow and at least one pair is not fast, the network is itself tight.

In the light traffic limit, the same benchmark would classify edges 1 and 3 as tight, edges 4 and 5 as slow, and edge 2 as fast. Under this classification, the first O/D pair would again be tight, but the second O/D pair would now be classified as slow (because all its paths contain a slow edge), so the network would no longer be tight. A moment's reflection shows that the reason for this is that the benchmark function  $c(x) = x$  is not well-suited for the second O/D pair. Instead, if we take the benchmark  $c(x) = 1$ , the first O/D pair would be classified as fast (because it has a fast path, namely edge 1) and the second pair would be classified as tight, so the network would now be tight. For a systematized version of this benchmark selection procedure, see the proof of Corollary 5.2 below.

With all this at hand, our next result states that if the costliest O/D pair in the network admits a tight path, selfish routing becomes asymptotically efficient in the limit:

**Theorem 5.1.** *Let  $\Gamma_M$  be a nonatomic routing game. If the network is tight in the limit as  $M \rightarrow \omega$ , then*

$$\lim_{M \rightarrow \omega} \text{PoA}(\Gamma_M) = 1. \quad (5.5)$$

In words, if (a) every O/D pair has a path which is not slow, and (b) the fastest path of the slowest O/D pair has a regularly varying cost, selfish routing becomes efficient in the limit. Motivated by the strong connection between Theorems 4.2 and 4.6, Theorem 5.1 has been stated in a way that does not discriminate between the light and heavy traffic regimes. The reason for this is to highlight the role of the tightness assumption: tightness at 0 guarantees efficiency in light traffic ( $\omega = 0$ ) while tightness at  $\infty$  guarantees efficiency in heavy traffic ( $\omega = \infty$ ).

Of course, both Theorems 4.2 and 4.6 follow as corollaries of Theorem 5.1 by taking respectively  $\omega = 0$  or  $\infty$  and specializing to a single O/D pair (in which case Eqs. (4.5) and (5.4) coincide). Other than that, however, the same caveats apply regarding the passage from light to heavy traffic: the classification of fast and slow edges could be reversed,



the equilibrium/socially optimum flows could be drastically different in the two regimes, etc. To illustrate all this, we proceed with some further corollaries of [Theorem 5.1](#) (which we prove in [Appendix B](#)):

**Corollary 5.2.** *If the network's costs are regularly varying at  $\omega$  and the (possibly infinite) limit*

$$\alpha_{e,e'} = \lim_{x \rightarrow \omega} c_e(x)/c_{e'}(x) \quad (5.6)$$

*exists for all  $e, e' \in \mathcal{E}$ , then  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \omega$ .*

*Proof.* Define a total preorder among the network's edges by setting  $e \leq e'$  if and only if  $\alpha_{e,e'} \leq 1$ . For each path  $p \in \mathcal{P}$ , choose a maximal element  $e_p$  of  $p$ , i.e., an edge  $e_p \in p$  such that  $e \leq e_p$  for all  $e \in p$ . Then, for each O/D pair  $i \in \mathcal{I}$ , choose a path  $p^i$  for which  $e_{p^i}$  is minimal, i.e.,  $e_{p^i} \leq e_{p'}$  for all  $p' \in \mathcal{P}^i$ . Finally, pick an O/D pair  $i \in \mathcal{I}$  such that  $e_{p^i}$  is maximal, i.e.,  $e_{p^i} \leq e_{p'}$  for all  $j \in \mathcal{I}$ . Setting  $e^* = e_{p^i}$  and taking  $c(x) \equiv c_{e^*}(x)$  as a benchmark, it is easy to verify that the network is tight at  $\omega$ , so [Theorem 5.1](#) applies.  $\square$

The proof of [Corollary 5.2](#) shows that the “comparison index”  $\alpha_{e,e'}$  induces a preference relation which refines the coarser classification of the network's edges into fast, slow and tight. Of course, this ordering could be reversed when passing from light to heavy traffic, but the existence thereof (along with regular variation) guarantees that the price of anarchy is asymptotically equal to 1 in both cases.

In addition to the above, [Corollary 5.2](#) also gives an alternative way to prove the following analogue of [Corollaries 4.3 and 4.8](#):

**Corollary 5.3.** *Suppose that the limit  $\lim_{x \rightarrow \omega} c_e(x)/x^{q_e}$  is finite and nonzero for some  $q_e \geq 0$  and all  $e \in \mathcal{E}$ . Then,  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \omega$ .*

*Proof.* Observe that

$$\frac{c_e(x)}{c_{e'}(x)} = \frac{c_e(x)}{x^{q_e}} \frac{x^{q_e}}{x^{q_{e'}}} \frac{x^{q_{e'}}}{c_{e'}(x)}, \quad (5.7)$$

so  $\lim_{x \rightarrow \omega} c_e(x)/c_{e'}(x)$  exists for all  $e \in \mathcal{E}$ . Our claim then follows from [Corollary 5.2](#).  $\square$

We thus obtain the following important corollary for polynomial cost functions:

**Corollary 5.4.** *In networks with polynomial costs,  $\text{PoA}(\Gamma_M) \rightarrow 1$  as  $M \rightarrow \omega$ .*

In words, [Corollary 5.4](#) yields the general principle that we stated in the introduction:

*In networks with polynomial costs,  
the price of anarchy becomes 1 under both light and heavy traffic.*

We find this principle particularly appealing as it indicates that the price of anarchy may attain high values *only* in an intermediate regime (where the traffic is neither light nor heavy).

## 6. NETWORKS WITH VARIABLE INFLOW RATES

An important assumption in the analysis of the previous section is that the relative inflow rate  $\lambda^i = m^i/M$  of each O/D pair  $i \in \mathcal{I}$  does not fluctuate in the limit – i.e., all pairs are assumed to generate a constant fraction of the overall traffic. In general however, this assumption need not hold: for instance, in an urban road network, central O/D pairs generate disproportionately more traffic during rush hour than peripheral, suburban destinations, so it is not reasonable to assume that traffic scales up maintaining a constant traffic ratio between different O/D pairs.



To understand the impact of this variability, consider two independent links,  $e_1$  and  $e_2$ , with corresponding cost functions  $c_1(x) = x$  and  $c_2(x) = x^2$ . Suppose further that these links are joining two uncoupled O/D pairs with inflow rates  $m^1$  and  $m^2$  and total inflow  $M = m^1 + m^2$ . If both inflows scale in the limit as  $\Theta(M)$ , the cost of the first pair will scale as  $\Theta(M)$  while that of the second pair will scale as  $\Theta(M^2)$ . As such, the leading contribution to congestion will be that of the first O/D pair in light traffic, and that of the second pair in heavy traffic. If, however, the inflow of the first pair scales as  $\Theta(M)$  but that of the second pair scales as  $\Theta(M^{1/3})$ , the induced costs will scale respectively as  $\Theta(M)$  and  $\Theta(M^{2/3})$ ; consequently, the most costly O/D pair will now be the second one in light traffic and the first one in heavy traffic.

This reversal of roles shows that the asymptotic behavior of the relative inflow rates  $\lambda^i = m^i/M$  could end up painting a completely different picture in the limit. In particular, if these relative rates oscillate wildly in the limit, the price of anarchy may exhibit a likewise irregular asymptotic behavior, even if the underlying network is well-behaved (for instance, even if it is tight; cf. [Example 6.2](#) below). As a result, special care must be taken to define and study the asymptotic regime in networks with variable traffic demands.

To that end, let  $\Gamma_n$  be a sequence of nonatomic routing games with total inflow  $M_n = \sum_{i \in \mathcal{I}} m_n^i$  induced by a sequence of inflow rates  $m_n^i$  for each O/D pair  $i \in \mathcal{I}$ . The light and heavy traffic limits are obviously recovered depending on whether the total inflow  $M_n$  converges respectively to  $\omega = 0$  or  $\omega = \infty$  as  $n \rightarrow \infty$ . However, the relative inflow rates  $\lambda_n^i = m_n^i/M_n$  could now exhibit very different behaviors as  $n \rightarrow \infty$ : in particular, as discussed above, the relative inflow of an O/D pair could oscillate – or even vanish – in the limit. To capture such phenomena, we introduce below the notion of *salience*:

**Definition 6.1.** Let  $\Gamma_n$  be a sequence of nonatomic routing games with relative inflow rates  $\lambda_n^i$ ,  $i \in \mathcal{I}$ . We say that a subset  $\mathcal{I}' \subseteq \mathcal{I}$  of O/D pairs is *salient* if

$$\liminf_{n \rightarrow \infty} \sum_{i \in \mathcal{I}'} \lambda_n^i > 0, \quad (6.1)$$

i.e., if the total fraction of the traffic generated by the O/D pairs in  $\mathcal{I}'$  is non-negligible in the limit.

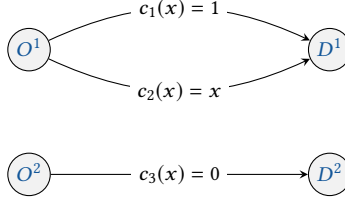
Obviously, if the sequence of relative inflow vectors  $\lambda_n = (\lambda_n^i)_{i \in \mathcal{I}}$  converges to a well-defined limit,  $\mathcal{I}'$  will be salient if and only if some O/D pair of  $\mathcal{I}'$  is itself salient – i.e., if and only if  $\liminf_{n \rightarrow \infty} \lambda_n^i > 0$  for some  $i \in \mathcal{I}'$ . However, if this is not the case, a set of O/D pairs may be salient even if none of its constituent pairs is salient: for instance, if there are two O/D pairs, “+” and “−”, with relative inflows  $\lambda_n^\pm = (1 \pm (-1)^n)/2$ , neither pair is salient but their union is (since  $\lambda_n^+ + \lambda_n^- = 1$  for all  $n$ ). Thus, the notion of salience does not rule out fluctuations in the relative inflows of *individual* O/D pairs; it only posits that the set of O/D pairs in question carries enough traffic in the limit.

Bearing all this in mind, our main result for networks with variable inflow rates is as follows:

**Theorem 6.2.** Let  $\Gamma_n$  be a sequence of nonatomic routing games with inflow rates  $m_n^i$  and total inflow  $M_n = \sum_{i \in \mathcal{I}} m_n^i$ . Suppose further that:

- (a) Traffic is either light or heavy in the limit, i.e.,  $\lim_{n \rightarrow \infty} M_n = \omega \in \{0, \infty\}$ .
- (b) Every O/D pair has a path which is not slow, i.e.,  $\alpha^i < \infty$  for all  $i \in \mathcal{I}$ .
- (c) The set of tight O/D pairs is salient, i.e.,  $\liminf_{n \rightarrow \infty} \sum_{i: \alpha^i > 0} \lambda_n^i > 0$ .

Then,  $\text{PoA}(\Gamma_n) \rightarrow 1$  as  $n \rightarrow \infty$ .



**Figure 4.** A simple network with two uncoupled O/D pairs.

Heuristically, [Condition \(a\)](#) above simply indicates the traffic regime under study (light or heavy), whereas [Condition \(b\)](#) guarantees that the network’s benchmark function correctly classifies the paths that are not too costly in each O/D pair. Finally, [Condition \(c\)](#) guarantees that tight O/D pairs are indeed relevant in terms of traffic, i.e., they account for a non-negligible fraction of the total inflow.

In view of all this, [Theorem 6.2](#) can be seen as a direct extension of our “fixed-rate” analysis in [Sections 4](#) and [5](#): indeed, in the case of constant (positive) relative inflows, salience boils down to asking that the network admits at least one tight O/D pair, so [Theorem 5.1](#) can be obtained as a special case of [Theorem 6.2](#). Below, we provide some further corollaries of [Theorem 6.2](#) along these lines:

**Corollary 6.3.** *If every O/D pair in the network is tight, then  $\text{PoA}(\Gamma_n) \rightarrow 1$ .*

**Corollary 6.4.** *If the network is tight and every O/D pair is salient, then  $\text{PoA}(\Gamma_n) \rightarrow 1$ .*

On the other hand, it is worth noting that if salience fails, we can draw no definitive conclusions for the price of anarchy. We illustrate the main reasons for this via two examples below:

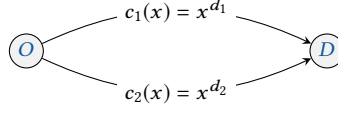
*Example 6.1 (Efficiency without salience).* Consider the simple network of [Fig. 4](#), where two “uncoupled” O/D pairs respectively encounter a standard Pigou network and an independent link with zero cost. In heavy traffic, the benchmark function  $c(x) = 1$  classifies the O/D pair  $i = 1$  as tight and the pair  $i = 2$  as fast, so the network is itself tight. Note also that the second O/D pair does not affect the network’s price of anarchy because it has a single routing option and its cost is identically equal to 0; however, it still affects the definition of relative inflows.

If we take the inflow sequence  $m_n^1 = \sqrt{n}$  and  $m_n^2 = n$ , we get  $M_n = n + \sqrt{n} \rightarrow \infty$  and  $\lambda_n^1 \rightarrow 0$  as  $n \rightarrow \infty$ , so the first O/D pair is not salient. Since the second O/D pair is not tight, [Condition \(c\)](#) fails; nevertheless, if we apply [Theorem 4.6](#) to the first O/D pair by itself, we obtain  $\lim_{M \rightarrow \infty} \text{PoA}(\Gamma_M) = 1$  (recall here that the second pair does not affect the network’s PoA). In other words, [Condition \(c\)](#) is not necessary for selfish routing to be asymptotically efficient.

*Example 6.2 (Inefficiency without salience).* Consider the same network as above but take the periodically oscillating inflow sequence

$$m_n^1 = \begin{cases} 1 & \text{for } n \text{ odd,} \\ 1 + 2n & \text{for } n \text{ even,} \end{cases} \quad \text{and} \quad m_n^2 = \begin{cases} 2n & \text{for } n \text{ odd,} \\ 0 & \text{for } n \text{ even.} \end{cases} \quad (6.2)$$

We then have  $M_n = 2n + 1 \rightarrow \infty$  and  $\liminf_{n \rightarrow \infty} \lambda_n^1 = 0$  so, again, [Condition \(c\)](#) fails (but in a different way). This time, whenever  $n$  is odd, the network’s price of anarchy is equal to that of a Pigou network with inflow 1 (because the second O/D pair is costless). Thus,



**Figure 5.** A two-link Pigou network with monomial costs.

for  $n$  odd, we get  $\text{PoA}(\Gamma_n) = 4/3$ , which is the worst-case value for networks with affine costs; as such, the conclusion of [Theorem 6.2](#) does not hold in general if we just drop the salience condition.

The above examples suggest that there is a finer mechanism at work which is not captured by the intersection of tightness and salience. At a high level, the crucial component of this mechanism seems to be that asymptotic efficiency is guaranteed if the network remains tight after suitably modifying the network's cost functions to take into account the scaling of the inflow of each O/D pair. However, getting an exact statement at this level of generality is fairly cumbersome, so we do not attempt it here.

## 7. RATE OF CONVERGENCE

The analysis of the previous sections provides a wide range of sufficient conditions guaranteeing that selfish routing becomes efficient in the limit; however, it does not provide any indication for the rate at which the network's price of anarchy converges to 1. In this section, we derive such rates (including subleading terms) for networks with polynomial costs of the form

$$c_e(x) = \sum_{k=q_e}^{d_e} c_{e,k} x^k, \quad (7.1)$$

where all coefficients are assumed nonnegative ( $c_{e,k} \geq 0$ ) and  $q_e$  and  $d_e$  respectively denote the smallest and largest powers present (so  $c_{e,q_e}, c_{e,d_e} > 0$ ); by convention, we also take  $q_e = \infty$  and  $d_e = 0$  when  $c_e(x) \equiv 0$ .<sup>3</sup> This model covers in particular the BPR “constant plus monomial” model (3.1) but also extends more easily to networks with more intricate cost functions.

In contrast to our qualitative analysis, the two traffic limits (light and heavy) exhibit different quantitative behaviors, so we treat them separately.

**7.1. The light traffic case.** We begin with the light traffic limit ( $\omega = 0$ ). To motivate our analysis, we start with a simple example of a Pigou network with monomial costs as shown in [Fig. 5](#): for  $d_1, d_2 > 0$ , the zero-flow travel time of both links is zero, so [Proposition 3.2](#) does not apply. Instead, as we show below, the network's price of anarchy decays to 1 following a power law:

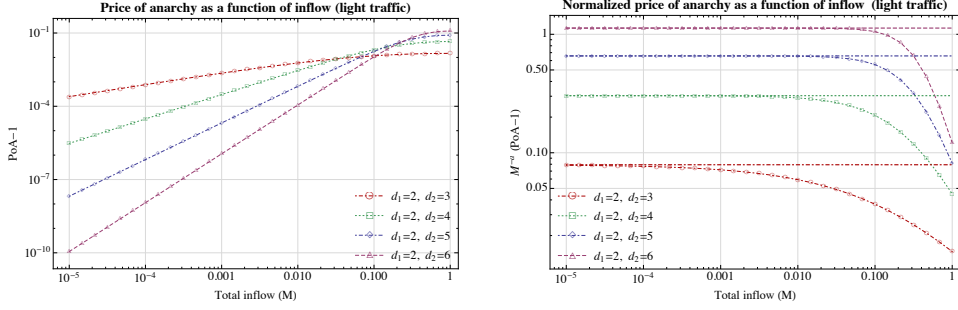
**Proposition 7.1.** *Consider a two-link parallel network with cost functions  $c_1(x) = x^{d_1}$  and  $c_2(x) = x^{d_2}$ ,  $0 < d_1 < d_2$ , and a single O/D pair with inflow  $M$ . Then*

$$\text{PoA}(\Gamma_M) = 1 + bM^a + o(M^a) \quad (7.2)$$

where

$$a = d_2/d_1 - 1, \quad (7.3)$$

<sup>3</sup>This follows the standard – if somewhat surprising at first – convention that  $\sup \emptyset = -\infty$ ,  $\inf \emptyset = \infty$ .



**Figure 6.** The rate of convergence of the price of anarchy in a lightly congested Pigou network as in Fig. 5. The figure on the left shows the price of anarchy as a function of the total traffic inflow  $M$  for different values of the exponents  $d_1$  and  $d_2$ . In the figure to the right, the price of anarchy has been rescaled by  $M^{-a}$  with  $a = d_2/d_1 - 1$ , showing that  $\text{PoA}(\Gamma_M) \sim 1 + bM^a$  for some  $b > 0$ ; the horizontal asymptotes correspond precisely to the expression (7.4) for  $b$ .

and

$$b = d_1 \left( \frac{1 + d_2}{1 + d_1} \right)^{1+1/d_1} - d_2 > 0. \quad (7.4)$$

In words, Proposition 7.1 shows that the rate of convergence of the price of anarchy is controlled by the ratio  $d_2/d_1$ : the largest the ratio of degrees, the fastest the decay of the price of anarchy (for a numerical illustration, see Fig. 6). This behavior is consistent with Proposition 3.2 which predicts that  $\text{PoA}(\Gamma_M) \equiv 1$  if  $M$  is small enough and  $d_1 = 0$ . Indeed, taking  $d_1 \rightarrow 0$  in (7.2) shows that  $\text{PoA}(\Gamma_M) = 1 + \mathcal{O}(M^a)$  for any  $a > 0$ , suggesting in turn that the rate of decay of  $\text{PoA}(\Gamma_M)$  is qualitatively different in this case.

Another case worth noting is when  $d_1 = d_2$ , i.e., when both links are equivalent in terms of performance. In this case,  $\text{PoA}(\Gamma_M)$  is identically equal to 1 for all values of  $M$  (Proposition 3.1 already guarantees as much when  $M$  is not large). However, (7.2) would seem to suggest that the price of anarchy can remain large as  $M \rightarrow 0$  (since  $a = 1 - 1 = 0$  when  $d_1 = d_2$ ). The solution of this apparent paradox is provided by looking at the multiplicative constant  $b$ : when  $d_1 = d_2$ , we also have  $b = 0$ , so the resulting contribution to the price of anarchy is 0 – not  $\Omega(1)$ .

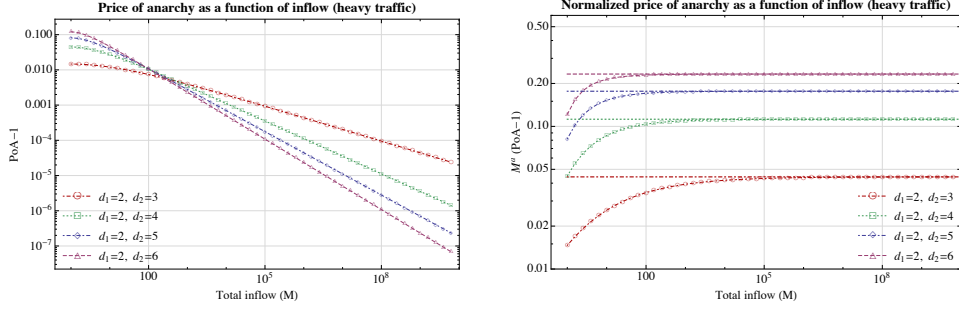
The above highlights the importance of the relative rate of decay of the network's edge costs as a function of the inflow. Since monomials with lower exponents are more costly in the low traffic limit, this rate is dominated by the smallest power in (7.1). Thus, motivated by the index machinery of Sections 4 and 5, we respectively define the *order* of an edge  $e \in \mathcal{E}$ , that of a path  $p \in \mathcal{P}$ , of an O/D pair  $i \in \mathcal{I}$ , and that of the network itself as

$$q_e = \min\{k : c_{e,k} > 0\} \quad (7.5a)$$

$$q_p = \min_{e \in p} q_e, \quad (7.5b)$$

$$q^i = \max_{p \in \mathcal{P}^i} q_p, \quad (7.5c)$$

$$q = \min_{i \in \mathcal{I}} q^i. \quad (7.5d)$$



**Figure 7.** The rate of convergence of the price of anarchy in a heavily congested Pigou network as in Fig. 5. The figure on the left shows the price of anarchy as a function of the total traffic inflow  $M$  for different values of the exponents  $d_1$  and  $d_2$ . In the figure to the right, the price of anarchy has been rescaled by  $M^a$  with  $a = 1 - d_1/d_2$ , showing that  $\text{PoA}(\Gamma_M) \sim 1 + b/M^a$  for some  $b > 0$ ; the horizontal asymptotes correspond precisely to the expression (7.11) for  $b$ .

In view of the above, the network is tight with respect to the benchmark function  $c(x) = x^q$ , and an edge  $e \in \mathcal{E}$  is fast when  $q_e > q$ , tight when  $q_e = q$ , and slow if  $q_e < q$ . We then denote the set of the network's slow edges as

$$\mathcal{E}_{\text{slow}} = \{e \in \mathcal{E} : q_e < q\}, \quad (7.6)$$

and we write

$$q_{\text{slow}} = \max_{e \in \mathcal{E}_{\text{slow}}} q_e \quad (7.7)$$

for the order of the fastest edge in  $\mathcal{E}_{\text{slow}}$  (again employing the standard convention that  $\max \emptyset = -\infty$ , so  $q_{\text{slow}} = -\infty$  when there are no slow edges).

Building on the intuition gained from Proposition 7.1, our main quantitative result for low traffic is that the network's PoA decays to 1 following a power law that depends only on the ratio between the order of the network ( $q$ ) and that of its fastest slow edge ( $q_{\text{slow}}$ ):

**Theorem 7.2.** *Let  $\Gamma_M$  be a nonatomic routing game with polynomial costs, total inflow  $M$ , and fixed relative inflows. Then, there exist non-negative constants  $K_1 \geq 0$  and  $K_a \geq 0$  such that*

$$\text{PoA}(\Gamma_M) \leq 1 + K_1 M + K_a M^a, \quad (7.8)$$

where  $a = q/q_{\text{slow}} - 1$  and  $K_a = 0$  whenever  $\mathcal{E}_{\text{slow}} = \emptyset$ .

Theorem 7.2 was stated for networks with fixed relative inflows for simplicity only: in Appendix C, we state and prove a more general result for networks with variable relative inflows as in Section 5. In terms of intuition, we only note here that Theorem 7.2 complements the insights gained from Propositions 3.1 and 3.2 in an important way: when there is no single “best path” under zero inflow, the network's price of anarchy is no longer identically equal to 1 for small inflows but instead behaves as a power law.

**7.2. The heavy traffic case.** We now turn to the heavy traffic limit ( $\omega = \infty$ ). As in the light traffic case, we start with a simple – but illuminating – example of a two-link Pigou network where precise results can be obtained:

**Proposition 7.3.** *Consider a two-link parallel network with cost functions  $c_1(x) = x^{d_1}$  and  $c_2(x) = x^{d_2}$ ,  $0 < d_1 < d_2$ , and a single O/D pair with inflow  $M$ . Then*

$$\text{PoA}(\Gamma_M) = 1 + bM^{-a} + o(M^{-a}) \quad (7.9)$$

where

$$a = 1 - d_1/d_2 \quad (7.10)$$

and

$$b = d_2 \left( \frac{1 + d_1}{1 + d_2} \right)^{1+1/d_2} - d_1 > 0. \quad (7.11)$$

For illustration purposes, we plotted in Fig. 7 the asymptotic behavior of the network's price of anarchy for different values of  $d_1$  and  $d_2$ . In the same vein as in the light traffic limit, two special cases that are of interest here are when  $d_2 = \infty$  and when  $d_1 = d_2$ . In the former ( $d_2 = \infty$ ), Eq. (7.10) gives  $a = 1$ , indicating a convergence rate of the order of  $\mathcal{O}(1/M)$ : since  $a < 1$  for all finite  $d_2$ , this is the best possible rate that can be achieved in the heavy traffic limit. For the latter ( $d_1 = d_2$ ), Eq. (7.10) gives  $a = 0$ , suggesting that the price of anarchy can remain large as  $M \rightarrow \infty$ . This seems to be inconsistent with the fact that the network's price of anarchy is identically equal to 1 when  $d_1 = d_2$  but a closer look reveals that the multiplicative constant  $b$  of (7.11) is also 0 when  $d_1 = d_2$ , thus reconciling the two results.

Now, even though the above result does not apply to more general networks with polynomial costs, it still highlights the main mechanism at play. Specifically, for large edge loads  $x$ , the dominant term in (7.1) is the one with highest degree  $d_e$ . As we've discussed before, this indicates a complete reversal of roles between light and heavy traffic: for  $d_1 < d_2$ ,  $x^{d_1}$  is *slower* than  $x^{d_2}$  when  $x \rightarrow 0$ , but *faster* when  $x \rightarrow \infty$ . Thus, with an obvious adaptation of what we did for light traffic, we define the *order* of an edge  $e \in \mathcal{E}$ , that of a path  $p \in \mathcal{P}$ , of an O/D pair  $i \in \mathcal{I}$ , and of the network itself as

$$d_e = \max\{k : c_{e,k} > 0\} \quad (7.12a)$$

$$d_p = \max_{e \in p} d_e, \quad (7.12b)$$

$$d^i = \min_{p \in \mathcal{P}^i} d_p, \quad (7.12c)$$

$$d = \max_{i \in \mathcal{I}} d^i. \quad (7.12d)$$

With all this at hand, we see that the network is tight with respect to the benchmark  $c(x) = x^d$ , so an edge  $e \in \mathcal{E}$  is fast when  $d_e < d$ , tight when  $d_e = d$ , and slow if  $d_e > d$ . The set of the network's slow edges is then denoted as

$$\mathcal{E}_{\text{slow}} = \{e \in \mathcal{E} : d_e > d\}, \quad (7.13)$$

and we write

$$d_{\text{slow}} = \min_{e \in \mathcal{E}_{\text{slow}}} d_e \quad (7.14)$$

for the order of the fastest edge in  $\mathcal{E}_{\text{slow}}$  (again employing the standard convention that  $\min \emptyset = \infty$ , so  $d_{\text{slow}} = \infty$  when there are no slow edges).

Mutatis mutandis, this definition is the same as in light traffic except for a reversal of the max/min operators. Our main result for heavily congested networks confirms this intuition:

**Theorem 7.4.** *Let  $\Gamma_M$  be a nonatomic routing game with polynomial costs, total inflow  $M$ , and fixed relative inflows. Then, there exist non-negative constants  $K_1 \geq 0$  and  $K_a \geq 0$  such that*

$$\text{PoA}(\Gamma_M) \leq 1 + \frac{K_1}{M} + \frac{K_a}{M^a}, \quad (7.15)$$

where  $a = 1 - d/d_{\text{slow}}$  and  $K_a = 0$  whenever  $\mathcal{E}_{\text{slow}} = \emptyset$ .

*Remark 7.1.* As in the light traffic case, if the costs are monomials of the same degree, then  $K_1 = K_a = 0$  and  $\text{PoA}(\Gamma_M) \equiv 1$  for all  $M > 0$ .

In words, given that  $a < 1$  when there is at least one slow edge in the network ( $a = 1$  and  $K_a = 0$  otherwise), [Theorem 7.4](#) states the network's price of anarchy converges to 1 as  $\mathcal{O}(1/M^a)$  with an  $\mathcal{O}(1/M)$  subleading term. In particular, in the presence of a single slow edge  $e$  with  $d_e > d$ , the convergence exponent  $a$  in (7.15) can become as small as  $1/(d+1)$ , thus pointing to a slower convergence rate in networks with routing costs of high degree and a small gap between the degree of tight and slow edges. On the other hand, if there are no slow edges we have  $\mathcal{E}_{\text{slow}} = \emptyset$  and we get an  $\mathcal{O}(1/M)$  rate of convergence.

On this issue, [Wu et al. \(2017\)](#) recently showed that if all the network's cost functions are of the BPR type  $c_e(x) = a_e + b_e x^d$  with the same degree  $d$ , then  $\text{PoA}(\Gamma_M) = 1 + \mathcal{O}(M^{-d})$  as  $M \rightarrow \infty$ . In this special case, the rate of convergence is faster than the prediction of [Theorem 7.4](#), a gap which points to a sharp discontinuity that occurs when all costs have the same degree. To see this in a concrete example, consider again the two-link Pigou network of [Fig. 5](#). By symmetry, if  $d_1 = d_2$ , the fraction of traffic routed on edge 1 at optimum and at equilibrium coincide

$$\tilde{y}_1 = y_1^* = \frac{1}{2} \quad \text{for all } M > 0, \quad (7.16)$$

implying in turn that  $\text{PoA}(\Gamma_M)$  is identically equal to 1. On the other hand, when  $d_1 < d_2$ , both fractions  $\tilde{y}_1$  and  $y_1^*$  converge to 1 as  $M \rightarrow \infty$ . [Proposition 7.3](#) shows that the rate of convergence of the price of anarchy in this case is exactly of order  $\Theta(1/M^a)$  and cannot be improved.

Put differently, [Proposition 7.3](#) shows that the slightest difference in edge degrees causes the rate of convergence of the price of anarchy to drop abruptly to  $\Theta(M^{-a})$ ; in fact, [Eq. \(7.11\)](#) even provides an explicit expression for the proportionality constant in the high congestion rate  $\Theta(M^{-a})$ . By this token, the bounds provided by [Theorem 7.4](#) are tight and cannot be improved in general, even in the class of two-link parallel networks with monomial costs.

## 8. DISCUSSION

Most of the literature on the price of anarchy – for congestion games and not only – has traditionally focused on worst-case upper bounds for different classes of networks, cost functions, and/or types of players. Several of these results have become milestones in the field and have had a significant impact in practical considerations for traffic networks. However, real-world situations involve a fixed network and traffic flows that are not necessarily close to these worst-case scenarios. Thus, in addition to determining *how* bad can selfish routing become, it is also important to determine *when* these cases are relevant.

Our goal in this paper was to provide an answer to this question by examining the behavior of the price of anarchy at each end of the congestion spectrum. Under fairly mild



assumptions (that always include networks with polynomial costs), we found that the PoA goes to 1 in both cases, independently of the network's topology, and even when there are multiple O/D pairs. What we find appealing about this result is that it is essentially independent of the underlying graph and/or the distribution of O/D pairs in the network. Especially in the heavy traffic limit, this means that selfish routing is not the real cause of increased delays: from a social planner's point of view, sophisticated tolling/rerouting schemes that target the optimum traffic assignment will not yield considerable gains over a "laissez-faire" approach where each traffic element takes the fastest available path.

#### APPENDIX A. PROOFS OF THE RESULTS OF SECTION 3

We start this appendix with the proofs of [Propositions 3.1](#) and [3.2](#). To that end, we first establish the following result:

**Lemma A.1.** *For sufficiently small  $M$ , equilibrium and optimum traffic allocations only employ paths in  $\mathcal{P}_{\min}^i = \arg \min_{p \in \mathcal{P}^i} c_p(0)$ .*

*Proof.* In a slight abuse of notation, let  $c_p(M) = \sum_{e \in p} c_e(M)$  denote the cost of the path  $p$  if all its edges carry load equal to the total inflow  $M$ . Clearly, if  $M$  is small enough, we have  $c_{p'}(0) > c_p(M)$  for all  $p \in \mathcal{P}_{\min}^i$  and all  $p' \in \mathcal{P}^i \setminus \mathcal{P}_{\min}^i$ . Hence, for an equilibrium flow  $f^*$ , we have

$$c_{p'}(f^*) \geq c_{p'}(0) > c_p(M) \geq c_p(f^*), \quad (\text{A.1})$$

implying in turn that  $f_{p'}^* = 0$ . Likewise, since an optimal flow  $\tilde{f}$  is an equilibrium for the marginal costs  $\tilde{c}_e(x) = c_e(x) + x c'_e(x)$  and  $\tilde{c}_{p'}(0) > \tilde{c}_p(M)$ , similar considerations show that an optimum flow profile cannot route any traffic along a path  $p' \in \mathcal{P}^i \setminus \mathcal{P}_{\min}^i$ .  $\square$

To proceed, it is more convenient to start with [Proposition 3.2](#):

*Proof of Proposition 3.2.* By [Lemma A.1](#), when  $M$  is small enough, both the equilibrium and the optimum must route the total inflow  $m^i$  along the unique path in  $\mathcal{P}_{\min}^i$ . Hence the equilibrium and optimal flows coincide and therefore the price of anarchy is equal to 1.  $\square$

With this result at hand, we have:

*Proof of Proposition 3.1.* If  $\mathcal{P}_{\min}^i$  is a singleton for all  $i \in \mathcal{I}$ , our claim follows from [Proposition 3.2](#). Otherwise, by [Lemma A.1](#), if  $M$  is small enough, for every  $i \in \mathcal{I}$ , only paths in  $\mathcal{P}_{\min}^i$  are used in equilibrium. Moreover, if  $p, p' \in \mathcal{P}_{\min}^i$ , then, for  $M$  small enough, we have

$$\sum_{e \in p} a_e + b_e(x_e^*)^d = \sum_{e \in p'} a_e + b_e(x_e^*)^d, \quad (\text{A.2})$$

and hence

$$\sum_{e \in p} b_e(x_e^*)^d = \sum_{e \in p'} b_e(x_e^*)^d. \quad (\text{A.3})$$

Again, by [Lemma A.1](#), if  $M$  is small enough, for every  $i \in \mathcal{I}$ , only paths in  $\mathcal{P}_{\min}^i$  are used at the optimum. If  $p, p' \in \mathcal{P}_{\min}^i$ , then, for  $M$  small enough, we have

$$\sum_{e \in p} a_e + (d+1)b_e \tilde{x}_e^d = \sum_{e \in p'} a_e + (d+1)b_e \tilde{x}_e^d, \quad (\text{A.4})$$

that is,

$$\sum_{e \in p} (d+1)b_e \tilde{x}_e^d = \sum_{e \in p'} (d+1)b_e \tilde{x}_e^d. \quad (\text{A.5})$$

Comparing (A.3) and (A.5), we see that the two equations are satisfied by the same loads. Therefore, for  $M$  small enough, there exist an equilibrium and an optimum having the same flows. Uniqueness of the equilibrium and optimum costs provides the result.  $\square$

We now present the proof of the counterexample with an oscillating PoA of Section 3.2:

*Proof of Proposition 3.3.* Since an unused edge has a cost of zero under (3.2), all three edges must be used at equilibrium. Hence, for a given value of the total inflow  $M = x_1 + x_2 + x_3$ , the load profile  $x = (x_1, x_2, x_3)$  is a Wardrop equilibrium if and only if  $c_1(x_1) = c_2(x_2) = c_3(x_3)$ . In that case, the normalized profile  $z = x/M$  satisfies

$$z_1^d \left[ 1 + \frac{1}{2} \sin(\log M z_1) \right] = z_2^d = z_3^d \left[ 1 + \frac{1}{2} \cos(\log M z_3) \right]. \quad (\text{A.6})$$

Likewise, after differentiating and rearranging, the conditions for the network's socially optimum flow are

$$z_1^d \left[ 1 + \frac{1}{2} \sin(\log M z_1) + \frac{1}{2(d+1)} \cos(\log M z_1) \right] = z_2^d = z_3^d \left[ 1 + \frac{1}{2} \cos(\log M z_3) - \frac{1}{2(d+1)} \sin(\log M z_3) \right]. \quad (\text{A.7})$$

We now show that Eqs. (A.6) and (A.7) never admit a common solution. Indeed, this can occur if and only if

$$\cos(\log M z_1) = 0 = \sin(\log M z_3), \quad (\text{A.8})$$

i.e., if and only if there exist integers  $k_1, k_3 \in \mathbb{Z}$  such that

$$\begin{aligned} \log M z_1 &= k_1 \pi + \frac{\pi}{2}, \\ \log M z_3 &= k_3 \pi. \end{aligned} \quad (\text{A.9})$$

This implies that  $\sin(\log M z_1) = \pm 1$  and  $\cos(\log M z_3) = \pm 1$ , leading to the following cases:

Case 1:  $\sin(\log M z_1) = 1$ ,  $\cos(\log M z_3) = 1$ . Substituting in (A.6) we get  $z_1^d = z_3^d$  so (A.9) gives

$$k_1 \pi + \frac{\pi}{2} = k_3 \pi. \quad (\text{A.10})$$

This gives  $k_3 - k_1 = 1/2$ , which cannot hold for integer values of  $k_1$  and  $k_3$ .

Case 2:  $\sin(\log M z_1) = 1$ ,  $\cos(\log M z_3) = -1$ . As above, from (A.6) we get  $3z_1^d = z_3^d$ , so (A.9) gives

$$\frac{1}{d} \log 3 + k_1 \pi + \frac{\pi}{2} = k_3 \pi. \quad (\text{A.11})$$

This yields  $\frac{\log 3}{\pi} = d(k_3 - k_1 - \frac{1}{2})$ , which again cannot hold for  $k_1, k_3, d \in \mathbb{Z}$ .

The remaining two cases lead to a contradiction in the same way, implying that the game's Wardrop equilibrium and socially optimum flows cannot coincide for any value of  $M$ . Since Eqs. (A.6) and (A.7) are periodic in  $\log M$ , it follows that the game's price of anarchy oscillates periodically at a logarithmic scale. Thus, focusing on the period  $1 \leq M \leq e^{2\pi}$ , we conclude that

$$\inf_{M > 0} \text{PoA}(\Gamma_M) = \min_{1 \leq M \leq e^{2\pi}} \text{PoA}(\Gamma_M) > 1, \quad (\text{A.12})$$

i.e., the Wardrop equilibria of the network in Fig. 1 remain strictly inefficient under both light and heavy traffic.  $\square$

## APPENDIX B. CONVERGENCE OF THE PRICE OF ANARCHY

We now prove [Theorem 6.2](#); [Theorems 4.2, 4.6](#) and [5.1](#) will then follow as special cases of this more general result. To that end, we begin with two auxiliary lemmas concerning the asymptotic behavior of regularly varying functions:

**Lemma B.1** ([Karamata, 1933](#)). *If  $g$  is regularly varying at  $\omega$ , there exists some  $\rho \in \mathbb{R}$  such that*

$$\lim_{t \rightarrow \omega} \frac{g(tx)}{g(t)} = x^\rho \quad \text{for all } x > 0. \quad (\text{B.1})$$

[Lemma B.1](#) is a classical result in the theory of regularly varying functions and gives rise to the term “ $\rho$ -regularly varying” for functions satisfying (B.1); for a proof, see, e.g., [Bingham et al. \(1989\)](#).

The second lemma is a more technical asymptotic comparison result allowing us to replace a  $\rho$ -regularly varying function by a monomial of degree  $\rho$  in the limit:

**Lemma B.2.** *Let  $\omega \in \{0, \infty\}$  and consider two functions  $f, g: (0, \infty) \rightarrow (0, \infty)$  such that:*

- (1)  *$f$  is nondecreasing.*
- (2)  *$g$  is  $\rho$ -regularly varying at  $\omega$  for some  $\rho > 0$ .*
- (3)  *$\lim_{x \rightarrow \omega} f(x)/g(x) = \alpha \in [0, \infty)$ .*

*If  $M_n \rightarrow \omega$  and  $z_n \rightarrow z \in [0, \infty)$ , then*

$$\lim_{n \rightarrow \infty} \frac{f(M_n z_n)}{g(M_n)} = \alpha z^\rho. \quad (\text{B.2})$$

*Proof.* We first consider the case  $\omega = \infty$ . If  $z > 0$ , the sequence  $x_n = M_n z_n$  diverges to infinity, so our claim follows from [Theorem 1.5.2](#) in [Bingham et al. \(1989\)](#) by writing

$$\frac{f(M_n z_n)}{g(M_n)} = \frac{f(x_n)}{g(x_n)} \cdot \frac{g(M_n z_n)}{g(M_n)} \rightarrow \alpha z^\rho. \quad (\text{B.3})$$

If  $z = 0$ , then, for all  $\varepsilon > 0$ , we have  $z_n \leq \varepsilon$  if  $n$  is sufficiently large. Then, using the monotonicity of  $f$  and the previous argument, we get

$$0 \leq \limsup_{n \rightarrow \infty} \frac{f(M_n z_n)}{g(M_n)} \leq \limsup_{n \rightarrow \infty} \frac{f(M_n \varepsilon)}{g(M_n)} = \alpha \varepsilon^\rho. \quad (\text{B.4})$$

Taking  $\varepsilon \rightarrow 0$ , we conclude that  $f(M_n z_n)/g(M_n) \rightarrow 0 = \alpha z^\rho$ , as claimed.

The case  $\omega = 0$  is even simpler. Indeed, we now have that  $x_n = M_n z_n$  tends to 0, so that the result follows using (B.3) and invoking [Theorem 1.5.2](#) in [Bingham et al. \(1989\)](#).  $\square$

Now, to proceed with the proof of [Theorem 6.2](#), we will require some additional notation. First, fix some inflow vector  $m = (m^i)_{i \in \mathcal{I}}$  with total inflow  $M = \sum_{i \in \mathcal{I}} m^i$  and relative inflows  $\lambda = (\lambda^i)_{i \in \mathcal{I}}$ . Instead of working directly with the flow variables  $f \in \mathcal{F}$ , it will be more convenient to introduce the *normalized traffic allocation* variables  $y^i = (y_p^i)_{p \in \mathcal{P}^i}$  defined as

$$y_p^i = f_p / m^i \quad \text{for all } p \in \mathcal{P}^i, i \in \mathcal{I}. \quad (\text{B.5})$$

We clearly have  $\sum_{p \in \mathcal{P}^i} y_p^i = 1$  for all  $i \in \mathcal{I}$ ; we will also write  $\mathcal{Y}^i = \Delta(\mathcal{P}^i)$  for the simplex of traffic allocations of  $i \in \mathcal{I}$  and  $\mathcal{Y} = \times_{i \in \mathcal{I}} \mathcal{Y}^i$  for the product thereof. Moreover, descending to the edge level, we define the *normalized load* induced by the  $i$ -th O/D pair on  $e \in \mathcal{E}$  as

$$z_e^i(y) = \sum_{p \in \mathcal{P}^i, p \ni e} y_p^i \quad (\text{B.6})$$

and we denote respectively the *normalized* and *total load* on edge  $e \in \mathcal{E}$  as

$$\zeta_e(y, \lambda) = \sum_{i \in \mathcal{I}} \lambda^i z_e^i(y) \quad \text{and} \quad x_e(y, m) = M \zeta_e(y, \lambda) = \sum_{i \in \mathcal{I}} m^i z_e^i(y). \quad (\text{B.7})$$

Finally, based on the index framework of Sections 4 and 5, we will respectively denote the set of the network's fast, tight and slow edges as

$$\mathcal{E}_{\text{fast}} = \{e \in \mathcal{E} : \alpha_e = 0\}, \quad (\text{B.8a})$$

$$\mathcal{E}_{\text{tight}} = \{e \in \mathcal{E} : 0 < \alpha_e < \infty\}, \quad (\text{B.8b})$$

$$\mathcal{E}_{\text{slow}} = \{e \in \mathcal{E} : \alpha_e = \infty\}, \quad (\text{B.8c})$$

and, in obvious notation, we will write e.g.,  $\mathcal{P}_{\text{slow}}$  for the set of the network's slow paths,  $\mathcal{I}_{\text{tight}}$  for the set of tight O/D pairs, etc.

The following asymptotic approximation result provides the heavy lifting for the proof of Theorem 6.2:

**Lemma B.3.** *Consider a network with nondecreasing cost functions  $g_e$ , with  $g_e(0) = 0$  for  $e \in \mathcal{E}$ , and suppose that it admits a benchmark function  $g$  at  $\omega$ , which is  $\rho$ -regularly varying with  $\rho > 0$ . Consider also a sequence of inflow vectors  $m_n = M_n \lambda_n$  such that:*

- a)  $M_n \rightarrow \omega$  and the vector of relative inflows  $\lambda_n$  converges to some  $\lambda \in \Delta(\mathcal{I})$ .
- b) Every O/D pair has a path which is not slow (relative to  $g$ ).
- c) There exists an O/D pair  $i \in \mathcal{I}$  which is tight (relative to  $g$ ) and has  $\lambda^i > 0$ .

Then, the optimal allocation problem

$$G_n = \min_{y \in \mathcal{Y}} \sum_{e \in \mathcal{E}} g_e(x_e(y, m_n)) \quad (\text{B.9})$$

satisfies

$$\lim_{n \rightarrow \infty} \frac{G_n}{g(M_n)} = V_\rho(\lambda), \quad (\text{B.10})$$

where  $V_\rho(\lambda) \in (0, \infty)$  is the solution value of the problem

$$V_\rho(\lambda) = \min_{y \in \mathcal{Y}} \sum_{e \in \mathcal{E}} \alpha_e \zeta_e(y, \lambda)^\rho \quad (\text{B.11})$$

and, by convention, we have set  $\alpha_e z_e^\rho = 0$  if  $\alpha_e = \infty$  and  $z_e = 0$ . Moreover, if  $\hat{y}_n$  is a sequence of optimal solutions of  $G_n$ , every limit point of  $\hat{y}_n$  solves  $V_\rho(\lambda)$ .

*Proof.* The arguments in the proof are similar to the line of reasoning in epi-convergence arguments as in Attouch (1984). To streamline the presentation, we break up the proof in five steps as follows:

*Step 1:*  $V_\rho(\lambda) < \infty$ . By Condition (b), each O/D pair admits a path that is not slow; therefore, routing all traffic through said path gives a finite value for the objective of (B.11), implying in turn that  $V_\rho(\lambda) < \infty$ . More precisely, for every  $i \in \mathcal{I}$ , take a traffic allocation  $y^i \in \mathcal{Y}^i$  that assigns zero weight to the slow paths  $\mathcal{P}_{\text{slow}}^i$  of  $i$ . Then, for every slow edge  $e \in \mathcal{E}_{\text{slow}}$ , we have  $z_e^i(y) = 0$  and, a fortiori,  $\zeta_e(y, \lambda) = 0$ ; hence

$$V_\rho(\lambda) \leq \sum_{e: \alpha_e < \infty} \alpha_e \zeta_e(y, \lambda)^\rho < \infty. \quad (\text{B.12})$$

*Step 2:*  $V_\rho(\lambda) > 0$ . By [Condition \(c\)](#), there exists a tight O/D pair  $i \in \mathcal{I}_{\text{tight}}$  such that  $\lambda^i > 0$ . For every  $y \in \mathcal{Y}$  we have  $\sum_{p \in \mathcal{P}^i} y_p = 1$ , so there exists some route  $p \in \mathcal{P}^i$  with  $y_p^i \geq 1/|\mathcal{P}^i|$ . This gives  $z_e^i(y) \geq 1/|\mathcal{P}^i|$  for all  $e \in p$ , and hence

$$\sum_{e \in \mathcal{E}} \alpha_e \zeta_e(y, \lambda)^\rho \geq \sum_{e \in \mathcal{E}} \alpha_e (\lambda^i z_e^i(y))^\rho \geq \sum_{e \in p} \alpha_e (\lambda^i / |\mathcal{P}^i|)^\rho \geq \alpha^i (\lambda^i / |\mathcal{P}^i|)^\rho > 0. \quad (\text{B.13})$$

Minimizing over  $y \in \mathcal{Y}$  then yields  $V_\rho(\lambda) > 0$ , as claimed.

*Step 3:*  $\limsup_{n \rightarrow \infty} G_n / g(M_n) \leq V_\rho(\lambda)$ . Fix an optimal solution  $\hat{y} \in \mathcal{Y}$  of [\(B.11\)](#). By the finiteness of  $V_\rho(\lambda)$ , we have  $\zeta_e(\hat{y}, \lambda) = 0$  for every slow edge  $e \in \mathcal{E}_{\text{slow}}$  (i.e., when  $\alpha_e = \infty$ ). If  $\lambda^i > 0$ , this implies that  $z_e^i(\hat{y}) = 0$ . Otherwise, if  $\lambda^i = 0$ , the objective function of [\(B.11\)](#) does not depend on  $y^i$ , so every  $y^i$  with  $z_e^i(y) = 0$  is also optimal. Hence we can choose the solution  $\hat{y}$  of [\(B.11\)](#) so that all traffic is routed along edges that are not slow.

Now, from optimality we have

$$\frac{G_n}{g(M_n)} \leq \frac{1}{g(M_n)} \sum_{e \in \mathcal{E}} g_e(M_n \zeta_e(\hat{y}, \lambda_n)). \quad (\text{B.14})$$

Using [Lemma B.2](#), for every non-slow edge  $e \in \mathcal{E} \setminus \mathcal{E}_{\text{slow}}$  (i.e.,  $\alpha_e < \infty$ ), we get

$$\lim_{n \rightarrow \infty} \frac{g_e(M_n \zeta_e(\hat{y}, \lambda_n))}{g(M_n)} = \alpha_e \zeta_e(\hat{y}, \lambda)^\rho. \quad (\text{B.15})$$

Otherwise, if  $e \in \mathcal{E}_{\text{slow}}$  is slow (i.e.,  $\alpha_e = \infty$ ), we have  $\zeta_e(\hat{y}, \lambda_n) = 0$ ; thus, since  $g_e(0) = 0$ , we get

$$\lim_{n \rightarrow \infty} \frac{g_e(M_n \zeta_e(\hat{y}, \lambda_n))}{g(M_n)} = \lim_{n \rightarrow \infty} \frac{g_e(0)}{g(M_n)} = 0 = \alpha_e \zeta_e(\hat{y}, \lambda)^\rho. \quad (\text{B.16})$$

Combining the previous three displayed equations, we obtain

$$\limsup_{n \rightarrow \infty} \frac{G_n}{g(M_n)} \leq \sum_{e \in \mathcal{E}} \lim_{n \rightarrow \infty} \frac{g_e(M_n \zeta_e(\hat{y}, \lambda_n))}{g(M_n)} = \sum_{e \in \mathcal{E}} \alpha_e \zeta_e(\hat{y}, \lambda)^\rho = V_\rho(\lambda). \quad (\text{B.17})$$

*Step 4:*  $\liminf_{n \rightarrow \infty} G_n / g(M_n) \geq V_\rho(\lambda)$ . Passing to a subsequence if necessary, we may assume that  $\liminf_{n \rightarrow \infty} G_n / g(M_n)$  is attained as a limit. Thus, letting  $\hat{y}_n$  be a sequence of solutions of  $G_n$ , and taking a further subsequence if necessary, we may assume that  $\hat{y}_n$  converges to some  $\hat{y} \in \mathcal{Y}$ . Then, ignoring the network's slow edges, we have

$$\frac{G_n}{g(M_n)} \geq \sum_{e: \alpha_e < \infty} \frac{g_e(M_n \zeta_e(\hat{y}_n, \lambda_n))}{g(M_n)}, \quad (\text{B.18})$$

and hence, by [Lemma B.2](#), we obtain

$$\liminf_{n \rightarrow \infty} \frac{G_n}{g(M_n)} \geq \sum_{e: \alpha_e < \infty} \alpha_e \zeta_e(\hat{y}, \lambda)^\rho. \quad (\text{B.19})$$

To proceed, we will show that  $\zeta_e(\hat{y}, \lambda) = 0$  for every slow edge. Indeed, if this were not the case, we could find some  $\varepsilon > 0$  such that  $\zeta_e(\hat{y}_n, \lambda_n) > \varepsilon$  for all sufficiently large  $n$ . With  $g_e$  nondecreasing, we then get

$$\frac{G_n}{g(M_n)} \geq \frac{g_e(M_n \zeta_e(\hat{y}_n, \lambda_n))}{g(M_n)} \geq \frac{g_e(M_n \varepsilon)}{g(M_n)} = \frac{g_e(M_n \varepsilon)}{g(M_n \varepsilon)} \frac{g(M_n \varepsilon)}{g(M_n)} \rightarrow \alpha_e \varepsilon^\rho = \infty, \quad (\text{B.20})$$

in contradiction to Steps 1 and 3 above. From all this, it follows that

$$\liminf_{n \rightarrow \infty} \frac{G_n}{g(M_n)} \geq \sum_{e \in \mathcal{E}} \alpha_e \zeta_e(\hat{y}, \lambda)^\rho \geq V_\rho(\lambda). \quad (\text{B.21})$$

*Step 5: Optimality of limit points.* As above, let  $\hat{y}_n$  be a sequence of optimal solutions of (B.9) and, by descending to a subsequence if necessary, assume that it converges to some  $\hat{y} \in \mathcal{Y}$ . From the previous steps we have  $G_n/g(M_n) \rightarrow V_\rho(\lambda)$  so, proceeding as in Step 4, we get

$$V_\rho(\lambda) = \lim_{n \rightarrow \infty} \frac{G_n}{g(M_n)} \geq \sum_{e \in \mathcal{E}} \alpha_e \zeta_e(\hat{y}, \lambda)^\rho \geq V_\rho(\lambda), \quad (\text{B.22})$$

showing that  $\hat{y}$  solves (B.11).  $\square$

Armed with Lemma B.3, we are finally in a position to prove Theorem 6.2:

*Proof of Theorem 6.2.* To begin with, express the objective function of (SO) in terms of the normalized flow variables  $y$  as

$$L_n(y) = \sum_{e \in \mathcal{E}} x_e(y, m_n) c_e(x_e(y, m_n)). \quad (\text{B.23})$$

Now, let  $y_n^*$ ,  $\tilde{y}_n$  be the normalized traffic allocation profiles of a Wardrop equilibrium and a socially optimum flow, respectively. Then, the network's price of anarchy may be expressed as

$$\text{PoA}(\Gamma_n) = \frac{\text{Eq}(\Gamma_n)}{\text{Opt}(\Gamma_n)} = \frac{L_n(y_n^*)}{L_n(\tilde{y}_n)}. \quad (\text{B.24})$$

Notice that  $\text{Opt}(\Gamma_n) > 0$  thanks to Assumptions (b) and (c).

In order to prove that  $\text{PoA}(\Gamma_n) \rightarrow 1$  it suffices to take a subsequence  $\Gamma_{n_k}$  realizing the  $\limsup_{n \rightarrow \infty} \text{PoA}(\Gamma_n)$  as a limit and to prove that  $\text{PoA}(\Gamma_{n_k}) \rightarrow 1$ . Relabeling indices and extracting a further subsequence if necessary, we may assume without loss of generality that: (a) the limit  $\lim_{n \rightarrow \infty} \text{PoA}(\Gamma_n)$  exists; (b) the sequence  $\lambda_n$  of relative inflows converges to some  $\lambda \in \Delta(\mathcal{I})$ ; and (c) the sequences  $y_n^*$  and  $\tilde{y}_n$  converge to some  $y^*$ ,  $\tilde{y} \in \mathcal{Y}$  respectively. With all this, we will use Lemma B.3 to derive the asymptotic behavior of  $\text{Opt}(\Gamma_n)$  and  $\text{Eq}(\Gamma_n)$ .

First, for  $\text{Opt}(\Gamma_n)$ , combining Lemma B.1 with Proposition 1.5.1 of Bingham et al. (1989) and the fact that the network's cost functions are nondecreasing, we immediately see that the network's benchmark function  $c$  is  $\beta$ -regularly varying for some  $\beta \geq 0$ . Then, letting  $g_e(x) = x c_e(x)$  and  $g(x) = x c(x)$ , we also get that  $g$  is  $\rho$ -regularly varying with  $\rho = 1 + \beta > 0$  and  $\lim_{x \rightarrow \omega} g_e(x)/g(x) = \lim_{x \rightarrow \omega} c_e(x)/c(x) = \alpha_e$ . This means that the hypotheses of Lemma B.3 are all satisfied, implying in turn that

$$L_n(\tilde{y}_n) = \text{Opt}(\Gamma_n) \sim V_\rho(\lambda) g(M_n) \quad \text{as } n \rightarrow \infty, \quad (\text{B.25})$$

with the notation " $f_n \sim g_n$ " meaning here that  $\lim_{n \rightarrow \infty} f_n/g_n = 1$ .

In view of this, and since  $0 < V_\rho(\lambda) < \infty$ , it remains to show that  $L_n(y_n^*) \sim V_\rho(\lambda) g(M_n)$ . To that end, we first analyze the asymptotic behavior of the convex minimization problem

$$W(\Gamma_n) = \min_{y \in \mathcal{Y}} \sum_{e \in \mathcal{E}} C_e(x_e(y, m_n)) \quad (\text{B.26})$$

by applying Lemma B.3 to the primitives  $C_e$  and  $C$  of  $c_e$  and  $c$  respectively. By a standard result (Bingham et al., 1989, Theorem 1.5.11),  $C$  is  $\rho$ -regularly varying with  $\rho = 1 + \beta$ ; moreover, by L'Hôpital's rule we also have  $\lim_{x \rightarrow \omega} C_e(x)/C(x) = \lim_{x \rightarrow \omega} c_e(x)/c(x) = \alpha_e$ . By Lemma B.3, it follows that  $W(\Gamma_n)/C(M_n) \rightarrow V_\rho(\lambda)$ . In addition, since the Wardrop equilibrium traffic allocations  $y_n^*$  are solutions of  $W(\Gamma_n)$ , the limit  $y^*$  of  $y_n^*$  is optimal for  $V_\rho(\lambda)$  by Lemma B.3.

Noting that  $x_e(y_n^*, m_n) = M_n \zeta_e(y_n^*, \lambda_n)$ , we obtain

$$\frac{L_n(y_n^*)}{g(M_n)} = \sum_{e \in \mathcal{E}} \frac{g_e(M_n \zeta_e(y_n^*, \lambda_n))}{g(M_n)}. \quad (\text{B.27})$$

By [Lemma B.2](#), we also have the following limit for every non-slow edge  $e \in \mathcal{E} \setminus \mathcal{E}_{\text{slow}}$ :

$$\frac{g_e(M_n \zeta_e(y_n^*, \lambda_n))}{g(M_n)} \rightarrow \alpha_e \zeta_e(y^*, \lambda)^\rho. \quad (\text{B.28})$$

To establish a similar limiting result when  $e$  is slow, we first claim that there exists a constant  $B \geq 0$  such that

$$g_e(M_n \zeta_e(y_n^*, \lambda_n)) \leq B \zeta_e(y_n^*, \lambda_n) g(M_n) \quad (\text{B.29})$$

This is trivially satisfied when  $\zeta_e(y_n^*, \lambda_n) = 0$ , so it suffices to consider the case  $\zeta_e(y_n^*, \lambda_n) > 0$ . The above inequality is then equivalent to asking that

$$c_e(M_n \zeta_e(y_n^*, \lambda_n)) \leq B c(M_n) \quad (\text{B.30})$$

Now,  $\zeta_e(y_n^*, \lambda_n) > 0$  implies that the edge  $e$  receives some equilibrium traffic from at least one O/D pair  $i \in \mathcal{I}$ , so it must belong to a path  $p \in \mathcal{P}^i$  with minimal cost. Thus, if we consider an alternative path  $p' \in \mathcal{P}^i$  all of whose edges are tight or fast, we have

$$c_e(M_n \zeta_e(y_n^*, \lambda_n)) \leq \sum_{e' \in p} c_{e'}(M_n \zeta_{e'}(y_n^*, \lambda_n)) \leq \sum_{e' \in p'} c_{e'}(M_n \zeta_{e'}(y_n^*, \lambda_n)). \quad (\text{B.31})$$

Using the trivial bound  $M_n \zeta_{e'}(y_n^*, \lambda_n) \leq M_n$ , we further get

$$c_e(M_n \zeta_e(y_n^*, \lambda_n)) \leq \sum_{e' \in p'} c_{e'}(M_n) \leq \sum_{e': \alpha_{e'} < \infty} c_{e'}(M_n). \quad (\text{B.32})$$

However, for every non-slow edge  $e' \in \mathcal{E} \setminus \mathcal{E}_{\text{slow}}$ , the sequence  $c_{e'}(M_n)/c(M_n)$  converges to  $\alpha_{e'}$  so we can find a constant  $B_{e'}$  such that  $c_{e'}(M_n)/c(M_n) \leq B_{e'}$  for all  $n \in \mathbb{N}$ ; consequently, (B.30) follows by taking  $B = \sum_{e': \alpha_{e'} < \infty} B_{e'}$ . Thus, given that  $y^*$  is optimal for  $V_\rho(\lambda)$ , we get  $\zeta_e(y_n^*, \lambda_n) \rightarrow \zeta_e(y^*, \lambda) = 0$ , and hence

$$\frac{g_e(M_n \zeta_e(y_n^*, \lambda_n))}{g(M_n)} \leq B \zeta_e(y_n^*, \lambda_n) \rightarrow 0 = \alpha_e \zeta_e(y^*, \lambda)^\rho. \quad (\text{B.33})$$

Combining (B.28), (B.33) and (B.27), we then get

$$L_n(y_n^*)/g(M_n) \rightarrow \sum_{e \in \mathcal{E}} \alpha_e \zeta_e(y^*, \lambda)^\rho = V_\rho(\lambda), \quad (\text{B.34})$$

as was to be shown.  $\square$

## APPENDIX C. SPEED OF CONVERGENCE

In this appendix, we provide the proofs of the results presented in [Section 7](#).

**C.1. Rates in the light traffic regime.** First we present the proof of [Proposition 7.1](#) on the light traffic rates in the case of a Pigou network.

*Proof of Proposition 7.1.* Let  $x$  denote the flow on edge  $e_1$ . At equilibrium, the costs on both edges must be equal so that  $(x^*)^{d_1} = (M - x^*)^{d_2}$ , which is equivalent to  $x^* + (x^*)^{d_1/d_2} = M$ . Since  $M$  tends to 0 it follows that  $x^*$  will be small and since  $d_1 < d_2$  the term  $(x^*)^{d_1/d_2}$  dominates the right hand side. Therefore

$$x^* = M^{d_2/d_1} (1 + o(1)) \quad (\text{C.1})$$



so the equilibrium cost  $\text{Eq}(\Gamma_M) = M \cdot c_1(x^*) = M \cdot c_2(M - x^*)$  scales as

$$\text{Eq}(\Gamma_M) = M \cdot \left[ M - M^{d_2/d_1}(1 + o(1)) \right]^{d_2} = M^{d_2+1} - d_2 M^{d_2+d_1/d_2} + o(M^{d_2+d_1/d_2}). \quad (\text{C.2})$$

Similarly, if  $\tilde{x}$  is the optimal flow on edge  $e_1$ , both edges have the same marginal cost

$$(1 + d_1)\tilde{x}^{d_1} = (1 + d_2)(M - \tilde{x})^{d_2}. \quad (\text{C.3})$$

Therefore, if we let

$$\rho = \left( \frac{1 + d_2}{1 + d_1} \right)^{1/d_2}, \quad (\text{C.4})$$

we get  $\rho\tilde{x} + \tilde{x}^{d_1/d_2} = \rho M$  as before, and hence

$$\tilde{x} = (\rho M)^{d_2/d_1}(1 + o(1)). \quad (\text{C.5})$$

It follows that the optimal cost scales as

$$\begin{aligned} \text{Opt}(\Gamma_M) &= \tilde{x} \cdot c_1(\tilde{x}) + (M - \tilde{x}) \cdot c_2(M - \tilde{x}) \\ &= \left[ (\rho M)^{d_2/d_1}(1 + o(1)) \right]^{d_1+1} + \left[ M - (\rho M)^{d_2/d_1}(1 + o(1)) \right]^{d_2+1} \\ &= (\rho M)^{d_2+d_2/d_1} + M^{d_2+1} - (d_2 + 1)M^{d_2}(\rho M)^{d_2/d_1} + o(M^{d_2+d_2/d_1}) \\ &= M^{d_2+1} - \rho^{d_2/d_1} \left[ (d_2 + 1) - \rho^{d_2} \right] M^{d_2+d_2/d_1} + o(M^{d_2+d_2/d_1}) \\ &= M^{d_2+1} - (b + d_2)M^{d_2+d_2/d_1} + o(M^{d_2+d_2/d_1}) \end{aligned} \quad (\text{C.6})$$

where the last equality follows from the identity  $\rho^{d_2/d_1} [(d_2 + 1) - \rho^{d_2}] = b + d_2$ .

Combining the previous expressions we get

$$\begin{aligned} \text{PoA}(\Gamma_M) &= \frac{\text{Eq}(\Gamma_M)}{\text{Opt}(\Gamma_M)} = \frac{\text{Opt}(\Gamma_M) + bM^{d_2+d_2/d_1} + o(M^{d_2+d_2/d_1})}{\text{Opt}(\Gamma_M)} \\ &= 1 + \frac{bM^{d_2+d_2/d_1} + o(M^{d_2+d_2/d_1})}{M^{d_2+1} + o(M^{d_2+1})} \\ &= 1 + bM^a + o(M^a). \end{aligned} \quad (\text{C.7})$$

To complete our proof, it remains to show that  $b > 0$ . After a small rearrangement, this is equivalent to establishing the inequality

$$\left( \frac{1 + d_2}{1 + d_1} \right)^{1+d_1} = \left( 1 + \frac{d_2 - d_1}{1 + d_1} \right)^{1+d_1} > \left( 1 + \frac{d_2 - d_1}{d_1} \right)^{d_1} = \left( \frac{d_2}{d_1} \right)^{d_1}, \quad (\text{C.8})$$

which itself follows from the fact that the function  $(1 + g/x)^x$  is increasing in  $x$  whenever  $g$  is positive.  $\square$

Now we prove the following more general version of [Theorem 7.2](#).

**Theorem C.1.** *Let  $\Gamma_n$  be a sequence of nonatomic routing games satisfying the assumptions of [Theorem 6.2](#), with  $M_n \rightarrow 0$ . Suppose further that the edge costs are polynomials as in [\(7.1\)](#), and let  $a = q/q_{\text{slow}} - 1$  with  $q$  and  $q_{\text{slow}}$  given by [\(7.5d\)](#) and [\(7.7\)](#) respectively. Then, there exist non-negative constants  $K_1 \geq 0$  and  $K_a \geq 0$  such that*

$$\text{PoA}(\Gamma_n) \leq 1 + K_1 M_n + K_a M_n^a \quad (\text{C.9})$$

with  $K_a = 0$  whenever  $\mathcal{E}_{\text{slow}} = \emptyset$ .

*Proof.* Let  $y_n^*, \tilde{y}_n \in \mathcal{Y}$  be an equilibrium and an optimum flow for  $\Gamma_n$  with induced edge flows  $x_{e,n}^* = M_n \zeta_e(y_n^*, \lambda_n)$  and  $\tilde{x}_{e,n} = M_n \zeta_e(\tilde{y}_n, \lambda_n)$ . The social cost of  $y_n^*$  can be estimated as

$$\begin{aligned}
L_n(y_n^*) &= \sum_{e \in \mathcal{E}} x_{e,n}^* c_e(x_{e,n}^*) = \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\
&= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \left[ \frac{q+1}{k+1} + \frac{k-q}{k+1} \right] c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\
&= (q+1) \sum_{e \in \mathcal{E}} C_e(x_{e,n}^*) + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{k-q}{k+1} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\
&\leq (q+1) \sum_{e \in \mathcal{E}} C_e(\tilde{x}_{e,n}) + \sum_{e \in \mathcal{E}} \sum_{k=q+1}^{d_e} \frac{k-q}{k+1} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \tag{C.10}
\end{aligned}$$

where  $C_e$  is the primitive of  $c_e$  and for the last inequality we used the fact that  $x_{e,n}^*$  minimizes the first sum, and in the double sum we dropped the negative terms with  $k \leq q$ . Now, the first sum in (C.10) can be further bounded as

$$\begin{aligned}
(q+1) \sum_{e \in \mathcal{E}} C_e(\tilde{x}_{e,n}) &= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{q+1}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\
&= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{q-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\
&\leq \text{Opt}(\Gamma_n) + \sum_{e \in \mathcal{E}_{\text{slow}}} \sum_{k=q_e}^{q-1} \frac{q-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1}
\end{aligned}$$

where we used the optimality of  $\tilde{x}_n$  in the first double sum, and we dropped the negative terms in the second. Note that in the latter only the slow edges with  $q_e < q$  are relevant. Combining these estimates we get

$$L_n(y_n^*) \leq \text{Opt}(\Gamma_n) + \sum_{e \in \mathcal{E}_{\text{slow}}} \sum_{k=q_e}^{q-1} \frac{q-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} + \sum_{e \in \mathcal{E}} \sum_{k=q+1}^{d_e} \frac{k-q}{k+1} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \tag{C.11}$$

Let us call  $L_n^I$  the first double sum in (C.11) and  $L_n^II$  the second. In order to bound  $L_n^II$  we assume that  $n$  is large enough so that  $M_n \leq 1$ . This assumption is done for convenience and it only affects the value of the constants  $K_1$  and  $K_a$ : by redefining them appropriately, the bound (C.9) will hold for all  $n$ . Then by setting

$$G = \sum_{e \in \mathcal{E}} \sum_{k=q+1}^{d_e} \frac{k-q}{k+1} c_{e,k}, \tag{C.12}$$

and noting that  $x_{e,n}^* \leq M_n$ , we get

$$L_n^II \leq \sum_{e \in \mathcal{E}} \sum_{k=q+1}^{d_e} \frac{k-q}{k+1} c_{e,k} \cdot M_n^{k+1} \leq G M_n^{q+2}. \tag{C.13}$$

In order to bound  $L_n^I$  we note that this term vanishes when  $\mathcal{E}_{\text{slow}}$  is empty. Otherwise, consider any edge  $e \in \mathcal{E}_{\text{slow}}$  that contributes to the sum with  $\tilde{x}_{e,n} > 0$ . We note that the

optimum flow  $\tilde{y}_n$  is an equilibrium for the marginal cost functions

$$\tilde{c}_e(x) = c_e(x) + x c'_e(x) = \sum_{k=q_e}^{d_e} (k+1)c_{e,k}x^k. \quad (\text{C.14})$$

The edge  $e$  must therefore belong to an optimal path  $p \in \mathcal{P}^i$  (w.r.t. the costs  $\tilde{c}_e(x)$ ) for some  $i \in \mathcal{I}$ . Hence, taking any alternative path  $p' \in \mathcal{P}^i$  which is not slow, i.e., with  $q_{e'} \geq q$  for all  $e' \in p'$ , and denoting

$$B = \sum_{e' \notin \mathcal{E}_{\text{slow}}} \sum_{k=q_{e'}}^{d_{e'}} (k+1)c_{e',k}, \quad (\text{C.15})$$

we get the bound (recall that  $M_n \leq 1$ )

$$\tilde{c}_e(\tilde{x}_{e,n}) \leq \sum_{e' \in p} \tilde{c}_{e'}(\tilde{x}_{e,n}) \leq \sum_{e' \in p'} \tilde{c}_{e'}(\tilde{x}_{e,n}) \leq \sum_{e' \in p'} \tilde{c}_{e'}(M_n) \leq BM_n^q. \quad (\text{C.16})$$

In particular, letting  $\tilde{c}_0 = \min_{e \in \mathcal{E}_{\text{slow}}} (q_e + 1)c_{e,q_e}$  we have

$$\tilde{c}_0 \cdot (\tilde{x}_{e,n})^{q_e} \leq (q_e + 1)c_{e,q_e} \cdot (\tilde{x}_{e,n})^{q_e} \leq \tilde{c}_e(\tilde{x}_{e,n}) \leq BM_n^q. \quad (\text{C.17})$$

Now, for  $n$  large we have  $BM_n^q/\tilde{c}_0 \leq 1$  and since  $q_e \leq q_{\text{slow}}$  we obtain  $\tilde{x}_{e,n} \leq (BM_n^q/\tilde{c}_0)^{1/q_{\text{slow}}}$ . Combining this latter bound with (C.16), and denoting  $D = B(B/\tilde{c}_0)^{1/q_{\text{slow}}} |\mathcal{E}_{\text{slow}}|$ , we deduce

$$\begin{aligned} L_n^1 &= \sum_{e \in \mathcal{E}_{\text{slow}}} \sum_{k=q_e}^{q-1} \frac{q-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\ &\leq \sum_{e \in \mathcal{E}_{\text{slow}}} \sum_{k=q_e}^{q-1} (k+1)c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\ &\leq \sum_{e \in \mathcal{E}_{\text{slow}}} \tilde{x}_{e,n} \tilde{c}_e(\tilde{x}_{e,n}) \\ &\leq \sum_{e \in \mathcal{E}_{\text{slow}}} (BM_n^q/\tilde{c}_0)^{1/q_{\text{slow}}} BM_n^q \\ &\leq DM_n^{q+q/q_{\text{slow}}}. \end{aligned} \quad (\text{C.18})$$

Plugging (C.18) and (C.13) into (C.11) we get

$$\text{PoA}(\Gamma_n) = \frac{L_n(y_n^*)}{\text{Opt}(\Gamma_n)} \leq \frac{\text{Opt}(\Gamma_n) + GM_n^{q+2} + DM_n^{q+q/q_{\text{slow}}}}{\text{Opt}(\Gamma_n)}. \quad (\text{C.19})$$

Now, if we set  $H = \min_{e \in \mathcal{E}} c_{e,q_e}$ , we have the following lower bound for the optimal cost

$$\text{Opt}(\Gamma_n) = \sum_{e \in \mathcal{E}} \tilde{x}_{e,n} \cdot c_e(\tilde{x}_{e,n}) \geq H \sum_{e \in \mathcal{E}} (\tilde{x}_{e,n})^{q_e+1}. \quad (\text{C.20})$$

We claim that the latter is of order at least  $O(M_n^{q+1})$ . Indeed, let us take  $\varepsilon > 0$  with  $\sum_{i \in \mathcal{I}_{\text{tight}}} \lambda_n^i \geq \varepsilon$  for sufficiently large  $n$ . For each  $n \in \mathbb{N}$  we may find  $i \in \mathcal{I}_{\text{tight}}$  such that  $\lambda_n^i \geq \varepsilon/|\mathcal{I}_{\text{tight}}|$  and, similarly, there exists a path  $p \in \mathcal{P}^i$  with  $\tilde{y}_{p,n} \geq 1/|\mathcal{P}^i| \geq 1/|\mathcal{P}|$ . Then, setting  $\kappa = 1/(|\mathcal{I}_{\text{tight}}| \times |\mathcal{P}|)$  we have  $\zeta_e(\tilde{y}_n, \lambda_n) \geq \kappa \varepsilon$  and therefore  $\tilde{x}_{e,n} \geq M_n \kappa \varepsilon$

for all  $e \in p$ . For  $n$  large we may assume that  $M_n \kappa \varepsilon \leq 1$  and, since the path  $p$  contains at least one edge  $e \in p$  with  $q_e \leq q$ , setting  $\bar{H} = H(\kappa \varepsilon)^{q+1}$  we get

$$\text{Opt}(\Gamma_n) \geq H(M_n \kappa \varepsilon)^{q+1} \geq \bar{H} M_n^{q+1}. \quad (\text{C.21})$$

This lower bound, combined with (C.19), yields (C.9) with  $K_1 = G/\bar{H}$  and  $K_a = D/\bar{H}$ . We conclude by noting that when  $\mathcal{E}_{\text{slow}} = \emptyset$  we have  $L_n^1 = 0$  and therefore we may take  $K_a = 0$ .  $\square$

**C.2. Rates in the heavy traffic regime.** We proceed with the proof of Proposition 7.3 on the heavy traffic rates in the case of a Pigou network.

*Proof of Proposition 7.3.* Let  $x$  denote the flow on edge  $e_2$ . At equilibrium, the costs on both edges must be equal so  $(x^*)^{d_2} = (M - x^*)^{d_1}$ , which is equivalent to  $x^* + (x^*)^{d_2/d_1} = M$ . It thus follows that

$$x^* = M^{d_1/d_2}(1 + o(1)), \quad (\text{C.22})$$

implying in turn that the equilibrium cost  $\text{Eq}(\Gamma_M) = M \cdot c_2(x^*) = M \cdot c_1(M - x^*)$  scales as

$$\text{Eq}(\Gamma_M) = M \cdot \left[ M - M^{d_1/d_2}(1 + o(1)) \right]^{d_1} = M^{1+d_1} - d_1 M^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2}). \quad (\text{C.23})$$

Similarly, if  $\tilde{x}$  is the optimal flow on edge  $e_2$ , both edges have the same marginal cost, namely

$$(1 + d_2)\tilde{x}^{d_2} = (1 + d_1)(M - \tilde{x})^{d_1}, \quad (\text{C.24})$$

and hence

$$\tilde{x} = (\theta M)^{d_1/d_2}(1 + o(1)), \quad (\text{C.25})$$

where we set

$$\theta = \left( \frac{1 + d_1}{1 + d_2} \right)^{1/d_1}. \quad (\text{C.26})$$

Therefore, the optimal social cost scales as

$$\begin{aligned} \text{Opt}(\Gamma_M) &= (M - \tilde{x}) \cdot c_1(M - \tilde{x}) + \tilde{x} \cdot c_2(\tilde{x}) \\ &= \left[ M - (\theta M)^{d_1/d_2}(1 + o(1)) \right]^{d_1+1} + \left[ (\theta M)^{d_1/d_2}(1 + o(1)) \right]^{d_2+1} \\ &= M^{d_1+1} - (d_1 + 1)M^{d_1}(\theta M)^{d_1/d_2} + (\theta M)^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2}) \\ &= M^{d_1+1} - \theta^{d_1/d_2}[(d_1 + 1) - \theta^{d_1}]M^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2}), \end{aligned} \quad (\text{C.27})$$

$$= M^{d_1+1} - (b + d_1)M^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2}), \quad (\text{C.28})$$

where  $b$  is defined as in (7.11).

Combining the previous expressions, we then get

$$\text{PoA}(\Gamma_M) = \frac{\text{Opt}(\Gamma_M) + bM^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2})}{\text{Opt}(\Gamma_M)} \quad (\text{C.29})$$

$$\begin{aligned} &= 1 + \frac{bM^{d_1+d_1/d_2} + o(M^{d_1+d_1/d_2})}{M^{d_1+1} + o(M^{d_1+1})} \\ &= 1 + bM^{-a} + o(M^{-a}), \end{aligned} \quad (\text{C.30})$$

which establishes the first part of Proposition 7.3. Finally, the positivity of  $b > 0$ , follows again from the fact that  $(1 - g/x)^x$  increased with  $x$  when  $g > 0$ .  $\square$

The following more general result subsumes Theorem 7.4.

**Theorem C.2.** *Let  $\Gamma_n$  be a sequence of nonatomic routing games satisfying the assumptions of Theorem 6.2, with  $M_n \rightarrow \infty$ . Suppose further that the edge costs are polynomials as in (7.1), and let  $a = 1 - d/d_{\text{slow}}$  with  $d$  and  $d_{\text{slow}}$  given by (7.12d) and (7.14). Then there exist non-negative constants  $K_1 \geq 0$  and  $K_a \geq 0$  such that*

$$\text{PoA}(\Gamma_n) \leq 1 + \frac{K_1}{M_n} + \frac{K_a}{M_n^a} \quad (\text{C.31})$$

with  $K_a = 0$  whenever  $\mathcal{E}_{\text{slow}} = \emptyset$ .

*Proof.* The proof follows a similar pattern as the one of Theorem C.1. Let again  $y_n^*, \tilde{y}_n \in \mathcal{Y}$  be an equilibrium and an optimum for  $\Gamma_n$ , respectively. Denote  $x_{e,n}^* = M_n \zeta_e(y_n^*, \lambda_n)$  and  $\tilde{x}_{e,n} = M_n \zeta_e(\tilde{y}_n, \lambda_n)$  the corresponding induced edge flows. As before, the social cost of  $y_n^*$  can be estimated as

$$\begin{aligned} L_n(y_n^*) &= \sum_{e \in \mathcal{E}} x_{e,n}^* c_e(x_{e,n}^*) = \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\ &= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \left[ \frac{d+1}{k+1} + \frac{k-d}{k+1} \right] c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\ &= (d+1) \sum_{e \in \mathcal{E}} C_e(x_{e,n}^*) + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{k-d}{k+1} c_{e,k} \cdot (x_{e,n}^*)^{k+1} \\ &\leq (d+1) \sum_{e \in \mathcal{E}} C_e(\tilde{x}_{e,n}) + \sum_{e \in \mathcal{E}_{\text{slow}}} x_{e,n}^* \cdot c_e(x_{e,n}^*), \end{aligned} \quad (\text{C.32})$$

where in the last inequality we used the fact that  $x_{e,n}^*$  minimizes the first sum, while in the double sum we dropped the edges  $e \notin \mathcal{E}_{\text{slow}}$  since  $(k-d)/(k+1) \leq 0$  for all  $k \leq d_e \leq d$ , and we used the inequality  $(k-d)/(k+1) \leq 1$  to bound the remaining terms  $e \in \mathcal{E}_{\text{slow}}$  by factoring out  $x_{e,n}^*$  and using the expression (7.1) for  $c_e(x)$ . Now, the first sum in (C.32) can be further bounded as

$$\begin{aligned} (d+1) \sum_{e \in \mathcal{E}} C_e(\tilde{x}_{e,n}) &= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{d+1}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\ &= \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d_e} \frac{d-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} \\ &\leq \text{Opt}(\Gamma_n) + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d-1} \frac{d-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1}, \end{aligned}$$

where in the inequality we used the optimality of  $\tilde{x}_n$  for the first sum and we dropped the negative terms in the second sum. Putting all this together we obtain the bound

$$L_n(y_n^*) \leq \text{Opt}(\Gamma_n) + \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d-1} \frac{d-k}{k+1} c_{e,k} \cdot (\tilde{x}_{e,n})^{k+1} + \sum_{e \in \mathcal{E}_{\text{slow}}} x_{e,n}^* \cdot c_e(x_{e,n}^*). \quad (\text{C.33})$$

Now, call  $L_n^{\text{I}}$  the first double sum and  $L_n^{\text{II}}$  the last sum in (C.33). In order to bound  $L_n^{\text{I}}$  we assume that  $n$  is large enough so that  $M_n \geq 1$ . Then, denoting

$$G = \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d-1} \frac{d-k}{k+1} c_{e,k} \quad (\text{C.34})$$

and using the fact that  $\tilde{x}_{e,n} \leq M_n$ , we can bound  $L_n^I$  as

$$L_n^I \leq \sum_{e \in \mathcal{E}} \sum_{k=q_e}^{d-1} \frac{d-k}{k+1} c_{e,k} \cdot M_n^{k+1} \leq GM_n^d. \quad (\text{C.35})$$

In order to bound  $L_n^{II}$  we note that this term vanishes whenever  $\mathcal{E}_{\text{slow}}$  is empty. Otherwise, consider any edge  $e \in \mathcal{E}_{\text{slow}}$  that contributes to the sum with  $x_{e,n}^* > 0$ . Since  $y_n^*$  is an equilibrium, the edge  $e$  must belong to a path  $p \in \mathcal{P}^i$  with minimal cost for some  $i \in \mathcal{I}$ . Hence, taking any alternative path  $p' \in \mathcal{P}^i$  which is not slow, and denoting

$$B = \sum_{e' \notin \mathcal{E}_{\text{slow}}} \sum_{k=q_{e'}}^{d_{e'}} c_{e',k}, \quad (\text{C.36})$$

we get the bound

$$c_e(x_{e,n}^*) \leq \sum_{e' \in p} c_{e'}(x_{e,n}^*) \leq \sum_{e' \in p'} c_{e'}(x_{e,n}^*) \leq \sum_{e' \in p'} c_{e'}(M_n) \leq BM_n^d. \quad (\text{C.37})$$

In particular, letting  $c_0 = \min_{e \in \mathcal{E}_{\text{slow}}} c_{e,d_e}$  we have

$$c_0 \cdot (x_{e,n}^*)^{d_e} \leq c_{e,d_e} \cdot (x_{e,n}^*)^{d_e} \leq c_e(x_{e,n}^*) \leq BM_n^d. \quad (\text{C.38})$$

Now, for  $n$  large we have  $BM_n^d/c_0 \geq 1$  and since  $d_e \geq d_{\text{slow}}$  we get  $x_{e,n}^* \leq (BM_n^d/c_0)^{1/d_{\text{slow}}}$ . Combining this latter bound with (C.37), and denoting  $D = B(B/c_0)^{1/d_{\text{slow}}} |\mathcal{E}_{\text{slow}}|$ , we deduce

$$L_n^{II} = \sum_{e \in \mathcal{E}_{\text{slow}}} x_{e,n}^* \cdot c_e(x_{e,n}^*) \leq \sum_{e \in \mathcal{E}_{\text{slow}}} \left( BM_n^d/c_0 \right)^{1/d_{\text{slow}}} BM_n^d \leq DM_n^{d+d_{\text{slow}}}. \quad (\text{C.39})$$

Plugging (C.35) and (C.39) into (C.33) we get

$$\text{PoA}(\Gamma_n) = \frac{L_n(y_n^*)}{\text{Opt}(\Gamma_n)} \leq \frac{\text{Opt}(\Gamma_n) + GM_n^d + DM_n^{d+d_{\text{slow}}}}{\text{Opt}(\Gamma_n)}. \quad (\text{C.40})$$

Now, if we set  $H = \min_{e \in \mathcal{E}} c_{e,d_e}$ , we have the following lower bound for the optimal cost

$$\text{Opt}(\Gamma_n) = \sum_{e \in \mathcal{E}} \tilde{x}_{e,n} \cdot c_e(\tilde{x}_{e,n}) \geq H \sum_{e \in \mathcal{E}} (\tilde{x}_{e,n})^{d_e+1}. \quad (\text{C.41})$$

We claim that the latter is of order at least  $O(M_n^{d+1})$ . Indeed, let us take  $\varepsilon > 0$  with  $\sum_{i \in \mathcal{I}_{\text{tight}}} \lambda_n^i \geq \varepsilon$  for sufficiently large  $n$ . For each  $n \in \mathbb{N}$  we may find  $i \in \mathcal{I}_{\text{tight}}$  such that  $\lambda_n^i \geq \varepsilon/|\mathcal{I}_{\text{tight}}|$  and, similarly, there exists a path  $p \in \mathcal{P}^i$  with  $\tilde{y}_{p,n} \geq 1/|\mathcal{P}^i| \geq 1/|\mathcal{P}|$ . Then, setting  $\kappa = 1/(|\mathcal{I}_{\text{tight}}| \times |\mathcal{P}|)$  we have  $\zeta_e(\tilde{y}_n, \lambda_n) \geq \kappa \varepsilon$  and therefore  $\tilde{x}_{e,n} \geq M_n \kappa \varepsilon$  for all  $e \in p$ . For  $n$  large we may assume that  $M_n \kappa \varepsilon \geq 1$  and, since the path  $p$  contains at least one edge  $e \in p$  with  $d_e \geq d$ , setting  $\bar{H} = H(\kappa \varepsilon)^{d+1}$  we get

$$\text{Opt}(\Gamma_n) \geq H(M_n \kappa \varepsilon)^{d_e+1} \geq \bar{H} M_n^{d+1}. \quad (\text{C.42})$$

This lower bound, combined with (C.40), yields (C.31) with  $K_1 = G/\bar{H}$  and  $K_a = D/\bar{H}$ . We conclude by noting that when  $\mathcal{E}_{\text{slow}} = \emptyset$  we have  $L_n^{II} = 0$  and therefore we may take  $K_a = 0$ .  $\square$

## REFERENCES

- Attouch, H. (1984). *Variational Convergence for Functions and Operators*. Pitman, Boston, MA.
- Beckmann, M. J., McGuire, C., and Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
- Bertsekas, D. P. and Gallager, R. (1992). *Data Networks*. Prentice Hall, Englewood Cliffs, NJ, 2 edition.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press, Cambridge, MA.
- Braess, D. (1968). über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12:258–268.
- Cole, R. and Tao, Y. (2016). Large market games with near optimal efficiency. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 791–808, New York, NY. ACM.
- Colini-Baldeschi, R., Cominetti, R., Mertikopoulos, P., and Scarsini, M. (2017). The asymptotic behavior of the price of anarchy. In *Web and Internet Economics*, pages 133–145.
- Colini-Baldeschi, R., Cominetti, R., and Scarsini, M. (2016). On the price of anarchy of highly congested nonatomic network games. In *Algorithmic Game Theory*, pages 117–128. Springer, Berlin.
- Correa, J. R., Schulz, A. S., and Stier-Moses, N. E. (2004). Selfish routing in capacitated networks. *Math. Oper. Res.*, 29(4):961–976.
- Correa, J. R., Schulz, A. S., and Stier-Moses, N. E. (2007). Fast, fair, and efficient flows in networks. *Oper. Res.*, 55(2):215–225.
- Correa, J. R., Schulz, A. S., and Stier-Moses, N. E. (2008). A geometric approach to the price of anarchy in nonatomic congestion games. *Games Econom. Behav.*, 64(2):457–469.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer, New York.
- Dumrauf, D. and Gairing, M. (2006). Price of anarchy for polynomial Wardrop games. In *WINE '06: Proceedings of the 2nd Conference on Web and Internet Economics*, pages 319–330. Springer, Berlin, Heidelberg.
- Feldman, M., Immorlica, N., Lucier, B., Roughgarden, T., and Syrgkanis, V. (2016). The price of anarchy in large games. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 963–976. ACM, New York.
- González Vayá, M., Grammatico, S., Andersson, G., and Lygeros, J. (2015). On the price of being selfish in large populations of plug-in electric vehicles. In *CDC '15: Proceedings of the 53rd IEEE Annual Conference on Decision and Control*, pages 6542–6547. IEEE.
- Jessen, A. H. and Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):171–192.
- Karamata, J. (1930). Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)*, 4:38–53.
- Karamata, J. (1933). Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Soc. Math. France*, 61:55–62.
- Koutsoupias, E. and Papadimitriou, C. H. (1999). Worst-case equilibria. In *STACS '99: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*.
- LeBlanc, L. J., Morlok, E. K., and Pierskalla, W. P. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, 9:309–318.
- Monnot, B., Benita, F., and Piliouras, G. (2017). How bad is selfish routing in practice? Technical report, arXiv:1703.01599v2.
- O'Hare, S. J., Connors, R. D., and Watling, D. P. (2016). Mechanisms that govern how the price of anarchy varies with travel demand. *Transportation Research Part B: Methodological*, 84:55–80.
- Papadimitriou, C. H. (2001). Algorithms, games, and the Internet. In *STOC '01: Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing*.
- Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan and Co., London, 1st edition.
- Resnick, S. I. (2007). *Heavy-Tail Phenomena*. Springer, New York.
- Roughgarden, T. (2003). The price of anarchy is independent of the network topology. *J. Comput. System Sci.*, 67(2):341–364.
- Roughgarden, T. (2007). Routing games. In *Algorithmic Game Theory*, pages 461–486. Cambridge Univ. Press, Cambridge.
- Roughgarden, T. and Tardos, E. (2002). How bad is selfish routing? *J. ACM*, 49(2):236–259.



- Roughgarden, T. and Tardos, E. (2004). Bounding the inefficiency of equilibria in nonatomic congestion games. *Games Econom. Behav.*, 47(2):389–403.
- Stidham, S. (2014). The price of anarchy for a network of queues in heavy traffic. In Pulat, P. S., Sarin, S. C., and Uzsoy, R., editors, *Essays in Production, Project Planning and Scheduling: A Festschrift in Honor of Salah Elmaghraby*, pages 91–121. Springer, Boston, MA.
- Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Part II*, volume 1, pages 325–378.
- Wu, Z., Möhring, R. H., and Chen, Y. (2017). Selfishness need not be bad. Technical report, arXiv:1712.07464v1.
- Youn, H., Gastner, M. T., and Jeong, H. (2008). Price of anarchy in transportation networks: Efficiency and optimality control. *Physical Review Letters*, 101(12):128701.